WO 2005/007870

5

20

25

30

COMPOSITIONS OF ORTHOGONAL LEUCYL-tRNA AND AMINOACYL-tRNA SYNTHETASE PAIRS AND USES THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority to and benefit of Provisional Patent Application USSN 60/485,451, filed July 7, 2003; and to Provisional Patent Application USSN 60/488,215, filed July 16, 2003, the disclosures of which are incorporated herein by reference in their entirety for all purposes.

STATEMENT AS TO RIGHTS TO INVENTIONS MADE UNDER FEDERALLY SPONSORED RESEARCH AND DEVELOPMENT

[0002] This invention was made with government support under Grant No. GM
62159 from the National Institutes of Health. The government may have certain rights to this invention.

FIELD OF THE INVENTION

[0003] The invention pertains to the field of translation biochemistry. The invention relates to methods for producing and compositions of orthogonal leucyl tRNAs, orthogonal leucyl aminoacyl-tRNA synthetases and pairs thereof. The invention also relates to methods of producing proteins in cells using such pairs and related compositions.

BACKGROUND OF THE INVENTION

[0004] The genetic code of every known organism, from bacteria to humans, encodes the same twenty common amino acids. Different combinations of the same twenty natural amino acids form proteins that carry out virtually all the complex processes of life, from photosynthesis to signal transduction and the immune response. In order to study and modify protein structure and function, scientists have attempted to manipulate both the genetic code and the amino acid sequence of protein. However, it has been difficult to remove the constraints imposed by the genetic code that limit proteins to twenty genetically encoded standard building blocks (with the rare exception of selenocysteine (see, e.g., A. Bock et al., (1991), Molecular Microbiology 5:515-20) and pyrrolysine (see, e.g., G. Srinivasan, et al., (2002), Science 296:1459-62).

V-165

20

25

30

. . . .

[0005]Some progress has been made to remove these constraints, although this progress has been limited and the ability to rationally control protein structure and function is still in its infancy. For example, chemists have developed methods and strategies to synthesize and manipulate the structures of small molecules (see, e.g., E. J. Corey, & X.-M. Cheng, The Logic of Chemical Synthesis (Wiley-Interscience, New York, 1995)). Total synthesis (see, e.g., B. Merrifield, (1986), Science 232:341-7 (1986)), and semi-synthetic methodologies (see, e.g., D. Y. Jackson et al., (1994) Science 266:243-7; and, P. E. Dawson, & S. B. Kent, (2000), Annual Review of Biochemistry 69:923-60), have made it possible to synthesize peptides and small proteins, but these methodologies have limited 10 utility with proteins over 10 kilo Daltons (kDa). Mutagenesis methods, though powerful, are restricted to a limited number of structural changes. In a number of cases, it has been possible to competitively incorporate close structural analogues of common amino acids throughout proteins. See, e.g., R. Furter, (1998), Protein Science 7:419-26; K. Kirshenbaum, et al., (2002), ChemBioChem 3:235-7; and, V. Doring et al., (2001), Science 15 292:501-4.

[0006] Early work demonstrated that the translational machinery of *E. coli* would accommodate amino acids similar in structure to the common twenty. *See*, Hortin, G., and Boime, I. (1983) Methods Enzymol. 96:777-784. This work was further extended by relaxing the specificity of endogenous *E. coli* synthetases so that they activate unnatural amino acids as well as their cognate natural amino acid. Moreover, it was shown that mutations in editing domains could also be used to extend the substrate scope of the endogenous synthetase. *See*, Doring, V., et al., (2001) Science 292:501-504. However, these strategies are limited to *recoding* the genetic code rather than *expanding* the genetic code and lead to varying degrees of substitution of one of the common twenty amino acids with an unnatural amino acid.

[0007] Later it was shown that unnatural amino acids could be site-specifically incorporated into proteins *in vitro* by the addition of chemically aminoacylated orthogonal amber suppressor tRNAs to an *in vitro* transcription/translation reaction. *See, e.g.*, Noren, C. J., et al. (1989) <u>Science</u> 244:182-188; Bain, J. D., et al., (1989) <u>J. Am. Chem. Soc.</u> 111:8013-8014; Dougherty, D. A. (2000) <u>Curr. Opin. Chem. Biol.</u> 4, 645-652; Cornish, V. W., et al. (1995) <u>Angew. Chem., Int. Ed.</u> 34:621-633; J. A. Ellman, et al., (1992), <u>Science</u> 255:197-200; and, D. Mendel, et al., (1995), Annual Review of Biophysics and

Biomolecular Structure 24:435-462. These studies show that the ribosome and translation factors are compatible with a large number of unnatural amino acids, even those with unusual structures. Unfortunately, the chemical aminoacylation of tRNAs is difficult, and the stoichiometric nature of this process severely limited the amount of protein that could be generated.

5

10

Unnatural amino acids have been microinjected into cells. For example, unnatural amino acids were introduced into the nicotinic acetylcholine receptor in Xenopus oocytes (e.g., M.W. Nowak, et al. (1998), In vivo incorporation of unnatural amino acids into ion channels in Xenopus oocyte expression system, Method Enzymol. 293:504-529) by microinjection of a chemically misacylated Tetrahymena thermophila tRNA (e.g., M.E. Saks, et al. (1996), An engineered Tetrahymena tRNAGln for in vivo incorporation of unnatural amino acids into proteins by nonsense suppression, J. Biol. Chem. 271:23169-23175), and the relevant mRNA. See, also, D.A. Dougherty (2000), Unnatural amino acids as probes of protein structure and function, Curr. Opin. Chem. Biol. 4:645-652.

Unfortunately, this methodology is limited to proteins in cells that can be microinicated.

Unfortunately, this methodology is limited to proteins in cells that can be microinjected, and, because the relevant tRNA is chemically acylated in vitro, and cannot be re-acylated, the yields of protein are very low.

[0009] To overcome these limitations, new components, e.g., orthogonal tRNAs, orthogonal aminoacyl-tRNA synthetases and pairs thereof, were added to the protein
20 biosynthetic machinery of the prokaryote Escherichia coli (E. coli) (see e.g., L. Wang, et al., (2001), Science 292:498-500), which allowed genetic encoding of unnatural amino acids in vivo. A number of new amino acids with novel chemical, physical or biological properties, including photoaffinity labels and photoisomerizable amino acids, photocrosslinking amino acids (see, e.g., Chin, J. W., et al. (2002) Proc. Natl. Acad. Sci. U.
25 S. A. 99:11020-11024; and, Chin, J. W., et al., (2002) J. Am. Chem. Soc. 124:9026-9027), keto amino acids (see, e.g., Wang, L., et al., (2003) Proc. Natl. Acad. Sci. U. S. A. 100:56-61), heavy atom containing amino acids, and glycosylated amino acids have been incorporated efficiently and with high fidelity into proteins in E. coli in response to, e.g., the amber codon (TAG), using this methodology.

30 [0010] Several other orthogonal pairs have been reported. Glutaminyl (see, e.g., Liu, D. R., and Schultz, P. G. (1999) Proc. Natl. Acad. Sci. U. S. A. 96:4780-4785), aspartyl (see, e.g., Pastrnak, M., et al., (2000) Helv. Chim. Acta 83:2277-2286), and tyrosyl (see,

e.g., Ohno, S., et al., (1998) J. Biochem. (Tokyo, Jpn.) 124:1065-1068; and, Kowal, A. K., et al., (2001) Proc. Natl. Acad. Sci. U. S. A. 98:2268-2273) systems derived from S. cerevisiae tRNAs and synthetases have been described for the potential incorporation of unnatural amino acids in E. coli. Systems derived from the E. coli glutaminyl (see, e.g., Kowal, A. K., et al., (2001) Proc. Natl. Acad. Sci. U. S. A. 98:2268-2273) and tyrosyl (see, e.g., Edwards, H., and Schimmel, P. (1990) Mol. Cell. Biol. 10:1633-1641) synthetase have been described for use in S. cerevisiae. The E. coli tyrosyl system has been used for the incorporation of 3-iodo-L-tyrosine in vivo, in mammalian cells. See, Sakamoto, K., et al., (2002) Nucleic Acids Res. 30:4692-4699. Typically, these systems have made use of the amber stop codon. To further expand the genetic code, there is a need to develop improved and/or additional components of the biosynthetic machinery, e.g., additional orthogonal tRNAs, orthogonal aminoacyl-tRNA synthetases, and/or unique codons. This invention fulfills these and other needs, as will be apparent upon review of the following disclosure.

SUMMARY OF THE INVENTION

- 15 [0011] To expand the genetic code, the invention provides compositions of and methods for producing orthogonal leucyl-tRNAs, orthogonal leucyl aminoacyl-tRNA synthetases and pairs thereof. These translational components can be used to incorporate a selected amino acid in a specific position in a growing polypeptide chain (during nucleic acid translation) in response to a selector codon.
- 20 [0012] Compositions of the invention include a composition comprising an orthogonal leucyl-tRNA (leucyl-O-tRNA), where the leucyl O-tRNA comprises an anticodon loop comprising a CU(X)_n XXXAA sequence, and comprises at least about a 25% suppression activity in presence of a cognate synthetase in response to a selector codon as compared to a comparable control (e.g., in the absence of the selector codon). In one embodiment, the selector codon is an amber codon, and the leucyl O-tRNA comprises a stem region comprising matched base pairs and a conserved discriminator base at position 73. This position is indicated in Figure 4, Panel A. In one aspect, the CU(X)_n XXXAA sequence comprises CUCUAAA sequence and n=0. In another aspect, the leucyl O-tRNA comprises a C:G base pair at position 3:70.
- 30 [0013] In one embodiment, the selector codon is a four-base codon and the leucyl O-tRNA comprises a first pair selected from U28:A42, G28:C42 and/or C28:G42, and a

second pair selected from G:49:C65 or C49:G65, where the numbering corresponds to that indicated in Figure 4, Panel A. In one aspect, the CU(X)_n XXXAA sequence comprises a CUUCCUAA sequence and n=1. In another aspect, the first pair is C28:G42 and the second pair is C49:G65. In one embodiment, the CU(X)_n XXXAA sequence comprises a CUUCAAA sequence and n=0, and the selector codon is an opal codon.

5

10

15

20

25

30

[0014] A composition comprising a leucyl O-tRNA can further include an orthogonal leucyl aminoacyl-tRNA synthetase (leucyl O-RS), where the leucyl O-RS preferentially aminoacylates the leucyl O-tRNA with a selected amino acid. In certain embodiments, a composition including a leucyl O-tRNA can further include a (e.g., in vitro or in vivo) translation system.

[0015] A composition of the invention also includes a cell (e.g., a non-eukaryotic cell (e.g., an *E. coli* cell), or a eukaryotic cell) comprising a translation system. The translation system includes an orthogonal leucyl-tRNA (leucyl-O-tRNA), where the leucyl-O-tRNA comprises at least about a 25% suppression activity in presence of a cognate synthetase in response to a selector codon as compared to a control lacking the selector codon; an orthogonal aminoacyl-leucyl-tRNA synthetase (leucyl-O-RS); and, a first selected amino acid. In these cells, the leucyl O-tRNA comprises an anticodon loop comprising a CU(X)_n XXXAA sequence and recognizes the first-selector codon and the leucyl O-RS preferentially aminoacylates the leucyl O-tRNA with the first selected amino acid. In some embodiments, the cell translation system comprises a leucyl-O-tRNA and cognate synthetase, or a conservative variant thereof, where these components are at least 50% as effective at suppressing a selector codon as a leucyl O-tRNA of SEQ ID NO: 3, 6, 7 or 12, in combination with a cognate synthetase.

[0016] In certain embodiments, the cell can further include an additional different O-tRNA/O-RS pair and a second selected amino acid, where the O-tRNA recognizes a second selector codon and the O-RS preferentially aminoacylates the O-tRNA with the second selected amino acid. In one embodiment, the cell further comprises a nucleic acid that comprises a polynucleotide that encodes a polypeptide of interest, where the polynucleotide comprises/encodes a selector codon that is recognized by the leucyl O-tRNA.

 $a_{i,j,k,l}$

5

10

15

20

30

[0017] In one embodiment, an *E. coli* cell includes an orthogonal leucyl-tRNA (leucyl-O-tRNA), where the leucyl-O-tRNA comprises at least about a 25% suppression activity in presence of a cognate synthetase in response to a selector codon as compared to a control lacking the selector codon; and an orthogonal leucyl aminoacyl-tRNA synthetase (leucyl-O-RS), where the O-RS preferentially aminoacylates the O-tRNA with a selected amino acid. The E. coli cell also includes the selected amino acid, and, a nucleic acid that comprises a polynucleotide that encodes a polypeptide of interest, where the polynucleotide comprises a selector codon that is recognized by the leucyl O-tRNA. In one example, the leucyl O-tRNA is derived from *Halobacterium sp NRC-1* and the leucyl O-RS is derived from *Methanobacterium thermoaautotropicum*.

[0018] In certain embodiments of the invention, a leucyl O-tRNA of the invention comprises or is encoded by a polynucleotide sequence as set forth in any one of SEQ ID NO.: 3, 6, 7 or 12, or a complementary polynucleotide sequence thereof. In some embodiments, the leucyl-O-tRNA and cognate synthetase, or a conservative variant thereof, are at least 50% as effective at suppressing a selector codon as a leucyl O-tRNA of SEQ ID NO: 3, 6, 7 or 12, in combination with a cognate synthetase. In the case of tRNA molecules, thymine (t) is, of course, replaced by uracil (u). In certain embodiments, a leucyl O-RS comprises an amino acid sequence as set forth in any one of SEQ ID NO.: 15 or 16, or a conservative variation thereof. In one embodiment, the leucyl O-RS or a portion thereof is encoded by a polynucleotide sequence as set forth in any one of SEQ ID NO.: 13 or 14, a conservative variant of SEQ ID NO: 13 or 14, or a complementary polynucleotide sequence thereof.

[0019] The leucyl O-tRNA and/or the leucyl O-RS of the invention can be derived from any of a variety of organisms (e.g., both eukaryotic and non-eukaryotic organisms).
 For example, the leucyl O-tRNA is derived from an archael tRNA (e.g., from Halobacterium sp NRC-1) and/or the leucyl O-RS is derived from a non-eukaryotic organism (e.g., Methanobacterium thermoaautotropicum).

[0020] Polynucleotides are also a feature of the invention. A polynucleotide of the invention includes a polynucleotide comprising a nucleotide sequence as set forth in any one of SEQ ID NO.: 1-2, 4-7, 12, and/or is complementary to or that encodes a polynucleotide sequence of the above. A polynucleotide of the invention also includes a nucleic acid that hybridizes to a polynucleotide described above, under highly stringent

conditions over substantially the entire length of the nucleic acid. A polynucleotide of the invention also includes a polynucleotide that is, e.g., at least 75%, at least 80%, at least 90%, at least 95%, at least 98% or more identical to that of a naturally occurring leucyl tRNA or a consensus sequence of multiple naturally occurring leucyl tRNAs, e.g., the laucyl tRNA of SEQ ID NO: 12, and comprises an anticodon loop comprising a CU(X)_n XXXAA sequence, an stem region lacking noncanonical base pairs and a conserved discriminator base at position 73. A polynucleotide of the invention also includes a polynucleotide that is, e.g., at least 75%, at least 80%, at least 90%, at least 95%, at least 98% or more identical to that of a naturally occurring leucyl tRNA and comprises an anticodon loop comprising a CUUCCUAA sequence, a first pair selected from T28:A42, G28:C42 and/or C28:G42, and a second pair selected from G:49:C65 or C49:G65, where the numbering corresponds to that indicated in Figure 4, Panel A. Polynucleotides that are, e.g., at least 80%, at least 90%, at least 95%, at least 98% or more identical to any of the above and/or a polynucleotide comprising a conservative variation of any the above or in Table 3 are also polynucleotides of the invention.

5

10

15

[0021] Vectors comprising or encoding a polynucleotide of the invention are also a feature of the invention. For example, a vector optionally includes any of: a plasmid, a cosmid, a phage, a virus, an expression vector, and/or the like. A cell comprising a vector of the invention is also a feature of the invention.

20 [0022] Methods of producing an orthogonal tRNA (O-tRNA), e.g., a leucyl OtRNA, are also a feature of the invention. An O-tRNA, e.g., a leucyl O-tRNA, produced by the method is also a feature of the invention. For example, a method includes mutating an anticodon loop on members of a pool of tRNAs (e.g., pool of leucyl tRNAs) to allow recognition of a selector codon, thereby providing a plurality of potential O-tRNAs; and 25 analyzing secondary structure of at least one member of the plurality potential O-tRNA to identify non-canonical base pairs in the secondary structure, and, optionally, mutating the non-canonical base pairs (e.g., mutating the non-canonical base pairs to canonical base pairs). In one embodiment, the non-canonical base pairs are located in stem region of the secondary structure. A population of cells of a first species, where the cells individually comprise at least one member of the plurality of potential O-tRNAs are subjected to a 30 negative selection, thereby eliminating cells that comprise a member of the plurality of potential O-tRNAs that is aminoacylated by an aminoacyl-tRNA synthetase (RS) that is

....

endogenous to the cell, and providing a pool of tRNAs that are orthogonal to the cell of the first species. In certain embodiments, the selector codon includes an amber codon, an opal codon, a four base codon, etc. The method can further include adding an additional sequence (CCA) to a 3' terminus of each of the pool of tRNAs and/or measuring suppression activity.

[0023] In one embodiment, the pool of tRNAs is obtained by aligning a plurality of tRNA sequences; determining a consensus sequence; and generating a library of mutant tRNAs using the consensus sequence, where the pool of tRNAs comprise the library of mutant tRNAs.

10 [0024] In certain embodiments, the subjecting step comprises a polynucleotide that encodes a negative selection marker. In one embodiment, the polynucleotide that encodes the negative selection marker comprises at least one selector codon. For example, a negative selection marker includes, but is not limited to, β-lactamase, β-galactosidase, and/or the like. In certain embodiments, the negative selection marker fluoresces or catalyzes a luminescent reaction in the presence of a suitable reactant. In another embodiment, a product of the negative selection marker is detected by fluorescence-activated cell sorting (FACS) or by luminescence. Optionally, the negative selection marker includes an affinity based screening marker. In certain embodiments, the subjecting step comprises growing the population of cells in the presence of a selective agent (e.g., an antibiotic, such as ampicillin).

In certain embodiments, the method further comprises subjecting to positive selection a second population of cells of the first species. The cells comprise a member of the pool of tRNAs that are orthogonal to the cell of the first species, a cognate aminoacyltRNA synthetase, and a positive selection marker. Cells are selected/screened for cells that comprise a member of the pool of tRNAs that is aminoacylated by the cognate aminoacyltRNA synthetase and that shows a desired response in the presence of the positive selection marker, thereby providing an O-tRNA.

25

30

[0026] Methods for identifying an orthogonal aminoacyl-tRNA synthetase (O-RS), e.g., a leucyl O-RS, for use with an O-tRNA, e.g., a leucyl O-tRNA, are also a feature of the invention. For example, a method includes subjecting to positive selection a population of cells of a first species, where the cells each comprise: 1) a member of a plurality of

aminoacyl-tRNA synthetases (RSs), where the plurality of RSs comprise mutant RSs, RSs derived from a species other than the first species or both mutant RSs and RSs derived from a species other than the first species; 2) the orthogonal tRNA (O-tRNA) (e.g., from a species other than the first species, from at least a second species, etc.); and 3) a polynucleotide that encodes a positive selection marker and comprises at least one selector codon. In one embodiment, the plurality of RSs comprises leucyl RSs. In certain embodiments, the O-tRNA comprises a leucyl O-tRNA (e.g., where leucyl O-tRNA includes at least about a 25% suppression activity in presence of a cognate synthetase in response to a selector codon as compared to a control lacking the cognate synthetase).

10 [0027]Cells are selected or screened for those that show an enhancement in suppression efficiency compared to cells lacking or having a reduced amount of the member of the plurality of RSs. These selected/screened cells comprise an active RS that aminoacylates the O-tRNA. The level of aminoacylation (in vitro or in vivo) by the active RS of a first set of tRNAs from the first species is compared to the level of aminoacylation 15 (in vitro or in vivo) by the active RS of a second set of tRNAs from a second species; where the level of aminoacylation is determined by a detectable substance (e.g., a labeled amino acid). The active RS that more efficiently aminoacylates the second set of tRNAs compared to the first set of tRNAs is selected, thereby providing the orthogonal aminoacyl-tRNA synthetase, e.g., leucyl O-RS, for use with the O-tRNA, e.g., the leucyl O-tRNA. An 20 orthogonal aminoacyl-tRNA synthetase identified by the method is also a feature of the invention.

[0028] Methods of producing a protein in a cell with a selected amino acid at a specified position are also a feature of the invention. For example, a method includes growing, in an appropriate medium, a cell, where the cell comprises a nucleic acid that comprises at least one selector codon and encodes a protein; and, providing the selected amino acid. The cell further comprises: an orthogonal leucyl-tRNA (leucyl-O-tRNA) that functions in the cell and recognizes the selector codon; and, an orthogonal leucyl aminoacyl-tRNA synthetase (leucyl O-RS) that preferentially aminoacylates the leucyl-O-tRNA with the selected amino acid. Typically, the leucyl-O-tRNA comprises at least about a 25% suppression activity in presence of a cognate synthetase in response to a selector codon as compared to a control lacking the cognate synthetase. A protein produced by this method is also a feature of the invention.

25

30

· 24 45

5

15

20

25

30

DEFINITIONS

ş

[0029] Before describing the invention in detail, it is to be understood that this invention is not limited to particular biological systems, which can, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting. As used in this specification and the appended claims, the singular forms "a", "an" and "the" include plural referents unless the content clearly dictates otherwise. Thus, for example, reference to "a cell" includes a combination of two or more cells; reference to "bacteria" includes mixtures of bacteria, and the like.

10 [0030] Unless defined herein and below in the reminder of the specification, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains.

[0031] Orthogonal leucyl-tRNA: As used herein, an orthogonal leucyl-tRNA (leucyl-O-tRNA) is a tRNA that is orthogonal to a translation system of interest, where the tRNA is: (1) identical or substantially similar to a naturally occurring leucyl tRNA, (2) derived from a naturally occurring leucyl tRNA by natural or artificial mutagenesis (3) derived by any process that takes a sequence of a wild-type or mutant leucyl tRNA sequence of (1) or (2) into account, (4) homologous to a wild-type or mutant leucyl tRNA; (5) homologous to any example tRNA that is designated as a substrate for a leucyl tRNA synthetase in Table 3, or (6) a conservative variant of any example tRNA that is designated as a substrate for a leucyl tRNA synthetase in Table 3. The leucyl tRNA can exist charged with anamino acid, or in an uncharged state. It is also to be understood that a "leucyl-O-tRNA" optionally is charged (aminoacylated) by a cognate synthetase with an amino acid other than leucine. Indeed, it will be appreciated that a leucyl-O-tRNA of the invention is advantageously used to insert essentially any amino acid, whether natural or artificial, into a growing polypeptide, during translation, in response to a selector codon.

Orthogonal leucyl amino acid synthetase: As used herein, an orthogonal leucyl amino acid synthetase (leucyl O-RS) is an enzyme that preferentially aminoacylates the leucyl-O-tRNA with an amino acid in a translation system of interest. The amino acid that the leucyl O-RS loads onto the leucyl O-tRNA can be any amino acid, whether natural or artificial, and is not limited herein. The synthetase is optionally the same as or homologous to a naturally occurring leucyl amino acid synthetase, or the same as or

homologous to a synthetase designated as a leucyl O-RS in Table 3. For example, the leucyl O-RS can be a conservative variant of a leucyl O-RS of Table 3, and/or can be at least 50%, 60%, 70%, 80%, 90%, 95%, 98%, 99% or more identical in sequence to a leucyl O-RS of Table 3.

[0033] Homologous: Proteins and/or protein sequences are "homologous" when they are derived, naturally or artificially, from a common ancestral protein or protein sequence. Similarly, nucleic acids and/or nucleic acid sequences are homologous when they are derived, naturally or artificially, from a common ancestral nucleic acid or nucleic acid sequence. For example, any naturally occurring nucleic acid can be modified by any available mutagenesis method to include one or more selector codon. When expressed, this 10 mutagenized nucleic acid encodes a polypeptide comprising one or more selected amino acid, e.g. unnatural amino acid. The mutation process can, of course, additionally alter one or more standard codon, thereby changing one or more standard amino acid in the resulting mutant protein as well. Homology is generally inferred from sequence similarity between 15 two or more nucleic acids or proteins (or sequences thereof). The precise percentage of similarity between sequences that is useful in establishing homology varies with the nucleic acid and protein at issue, but as little as 25% sequence similarity is routinely used to establish homology. Higher levels of sequence similarity, e.g., 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, or 99% or more, can also be used to establish homology. Methods for 20 determining sequence similarity percentages (e.g., BLASTP and BLASTN using default parameters) are described herein and are generally available.

Orthogonal: As used herein, the term "orthogonal" refers to a molecule (e.g., an orthogonal tRNA (O-tRNA) and/or an orthogonal aminoacyl tRNA synthetase (O-RS)) that functions with endogenous components of a cell with reduced efficiency as compared to a corresponding molecule that is endogenous to the cell or translation system, or that fails to function with endogenous components of the cell. In the context of tRNAs and aminoacyl-tRNA synthetases, orthogonal refers to an inability or reduced efficiency, e.g., less than 20 % efficiency, less than 10 % efficiency, less than 5 % efficiency, or less than 1% efficiency, of an orthogonal tRNA to function with an endogenous tRNA synthetase compared to the ability of an endogenous tRNA synthetase to function with an endogenous tRNA compared to the ability of an endogenous tRNA synthetase to function with the

25

30

endogenous tRNA. The orthogonal molecule lacks a functionally normal endogenous complementary molecule in the cell. For example, an orthogonal tRNA in a cell is aminoacylated by any endogenous RS of the cell with reduced or even undetectable efficiency, when compared to aminoacylation of an endogenous tRNA by the endogenous RS. In another example, an orthogonal RS aminoacylates any endogenous tRNA in a cell of interest with reduced or even undetectable efficiency, as compared to aminoacylation of the endogenous tRNA by an endogenous RS. A second orthogonal molecule can be introduced into the cell that functions with the first orthogonal molecule. For example, an orthogonal tRNA/RS pair includes introduced complementary components that function together in the cell with an efficiency (e.g., 45 % efficiency, 50% efficiency, 60% efficiency, 70% efficiency, 75% efficiency, 80% efficiency, 90% efficiency, 95% efficiency, or 99% or more efficiency) as compared to that of a control, e.g., a corresponding tRNA/RS endogenous pair, or an active orthogonal pair (e.g., a tyrosyl orthogonal tRNA/RS pair).

10

15

20

25

30

[0035] Cognate: The term "cognate" refers to components that function together, e.g., a leucyl tRNA and a leucyl aminoacyl-tRNA synthetase. The components can also be referred to as being complementary.

[0036] Preferentially aminoacylates: The term "preferentially aminoacylates" refers to an efficiency, e.g., 70 % efficient, 75 % efficient, 85% efficient, 90% efficient, 95 % efficient, or 99% or more efficient, at which an O-RS aminoacylates an O-tRNA with a selected amino acid, e.g., an unnatural amino acid, as compared to the O-RS aminoacylating a naturally occurring tRNA or a starting material used to generate the O-tRNA.

[0037] Selector codon: The term "selector codon" refers to codons recognized by the O-tRNA in the translation process and not recognized by an endogenous tRNA. The O-tRNA anticodon loop recognizes the selector codon on the mRNA and incorporates its amino acid, e.g., a selected amino acid, such as an unnatural amino acid, at this site in the polypeptide. Selector codons can include, e.g., nonsense codons, such as, stop codons, e.g., amber, ochre, and opal codons; four or more base codons; rare codons; codons derived from natural or unnatural base pairs and/or the like.

[0038] <u>Suppressor tRNA</u>: A suppressor tRNA is a tRNA that alters the reading of a messenger RNA (mRNA) in a given translation system, e.g., by providing a mechanism for incorporating an amino acid into a polypeptide chain in response to a selector codon. For

example, a suppressor tRNA can read through, e.g., a stop codon, a four base codon, or a rare codon.

[0039] Suppression activity: As used herein, the term "suppression activity" refers, in general, to the ability of a tRNA (e.g., a suppressor tRNA) to allow translational read-through of a codon (e.g. a selector codon that is an amber codon or a 4-or-more base codon) that would otherwise result in the termination of translation or mistranslation (e.g., frame-shifting). Suppression activity of a suppressor tRNA can be expressed as a percentage of translational read-through observed compared to a second suppressor tRNA, or as compared to a control system, e.g., a control system lacking an O-RS.

10

15

20

25

30

[0040] The present invention provides various means by which suppression activity can be quantitated. Percent suppression of a particular OtRNA and ORS against a selector codon (e.g., an amber codon) of interest refers to the percentage of activity of a given expressed test marker (e.g., LacZ), that includes a selector codon, in a nucleic acid encoding the expressed test marker, in a translation system of interest, where the translation system of interest includes an O-RS and an O-tRNA, as compared to a positive control construct, where the positive control lacks the O-tRNA, the O-RS and the selector codon. Thus, for example, if an active positive control marker construct that lacks a selector codon has an observed activity of X in a given translation system, in units relevant to the marker assay at issue, then percent suppression of a test construct comprising the selector codon is the percentage of X that the test marker construct displays under essentially the same environmental conditions as the positive control marker was expressed under, except that the test marker construct is expressed in a translation system that also includes the O-tRNA and the O-RS. Typically, the translation system expressing the test marker also includes an amino acid that is recognized by the O-RS and O-tRNA. Optionally, the percent suppression measurement can be refined by comparison of the test marker to a "background" or "negative" control marker construct, which includes the same selector codon as the test marker, but in a system that does not include the O-tRNA, O-RS and/or relevant amino acid recognized by the O-tRNA and/or O-RS. This negative control is useful in normalizing percent suppression measurements to account for background signal effects from the marker in the translation system of interest.

[0041] Suppression efficiency can be determined by any of a number of assays known in the art. For example, a β -galactosidase reporter assay can be used, e.g., a

Sec. 15.

10

15

25

derivatized lacZ plasmid (where the construct has a selector codon in the lacZ nucleic acid sequence) is introduced into cells from an appropriate organism (e.g., an organism where the orthogonal components can be used) along with plasmid comprising an O-tRNA of the invention. A cognate synthetase can also be introduced (either as a polypeptide or a polynucleotide that encodes the cognate synthetase when expressed). The cells are grown in media to a desired density, e.g., to an OD₆₀₀ of about 0.5, and β -galactosidase assays are performed, e.g., using the BetaFluorTM β -Galactosidase Assay Kit (Novagen). Percent suppression can be calculated as the percentage of activity for a sample relative to a comparable control, e.g., the value observed from the derivatived lacZ construct, where the construct has a corresponding sense codon at desired position rather than a selector codon.

- Translation system: The term "translation system" refers to the components that incorporate an amino acid into a growing polypeptide chain (protein). Components of a translation system can include, e.g., ribosomes, tRNAs, synthetases, mRNA and the like. The O-tRNA and/or O-RS of the invention can be added to or be a part of an *in vitro* or *in vivo* translation system, e.g., in a non-eukaryotic cell, e.g., a bacterium (such as *E coli*), or in a eukaryotic cell, e.g., a yeast cell, a mammalian cell, a plant cell, an algae cell, a fungus cell, an insect cell, and/or the like.
- [0043] Selected amino acid: The term "selected amino acid" refers to any desired naturally occurring amino acid or unnatural amino acid. As used herein, the term "unnatural amino acid" refers to any amino acid, modified amino acid, and/or amino acid analogue that is not one of the 20 common naturally occurring amino acids or seleno cysteine or pyrolysine.
 - [0044] Derived from: As used herein, the term "derived from" refers to a component that is isolated from or made using a specified molecule or organism, or information from the specified molecule or organism.
 - Positive selection or screening marker: As used herein, the term "positive selection or screening marker" refers to a marker that, when present, e.g., expressed, activated, or the like, results in identification of a cell with the positive selection marker from those without the positive selection marker.
- 30 [0046] <u>Negative selection or screening marker</u>: As used herein, the term "negative selection or screening marker" refers to a marker that, when present, e.g., expressed,

activated or the like, allows identification of a cell that does not possess a specified property (e.g., as compared to a cell that does possess the property).

[0047] Reporter: As used herein, the term "reporter" refers to a component that can be used to identify and/or select target components of a system of interest. For example, a

- 5 reporter can include a protein, e.g., an enzyme, that confers antibiotic resistance or sensitivity (e.g., β-lactamase, chloramphenicol acetyltransferase (CAT), and the like), a fluorescent screening marker (e.g., green fluorescent protein (e.g., (GFP), YFP, EGFP, RFP), a luminescent marker (e.g., a firefly luciferase protein), an affinity based screening marker, or positive or negative selectable marker genes such as lacZ, β-gal/lacZ (β-galactosidase), Adh (alcohol dehydrogenase), his3, ura3, leu2, lys2, or the like.
 - [0048] <u>Eukaryote</u>: As used herein, the term "eukaryote" refers to organisms belonging to the phylogenetic domain Eucarya, such as animals (e.g., mammals, insects, reptiles, birds, etc.), ciliates, plants (e.g., monocots, dicots, algae, etc.), fungi, yeasts, flagellates, microsporidia, protists, etc.
- Non-eukaryote: As used herein, the term "non-eukaryote" refers to non-eukaryotic organisms. For example, a non-eukaryotic organism can belong to the Eubacteria (e.g., Escherichia coli, Thermus thermophilus, Bacillus stearothermophilus, etc.) phylogenetic domain, or the Archaea (e.g., Methanococcus jannaschii, Methanobacterium thermoautotrophicum, Halobacterium such as Haloferax volcanii and Halobacterium species NRC-1, Archaeoglobus fulgidus, Pyrococcus furiosus, Pyrococcus horikoshii, Aeuropyrum pernix, etc.) phylogenetic domains.
 - translation component such as an O-tRNA or O-RS refers to a translation component that has a substantailly similar activity as the component on which the conservative variant is similar to, e.g., an O-tRNA or O-RS, but has variations in the sequence as compared to the base component. For example, an O-RS will aminoacylate a complementary O-tRNA or a conservative variant O-tRNA with a selected amino acid, e.g., an unnatural amino acid, although the O-tRNA and the conservative variant O-tRNA do not have the same sequence. The conservative variant can have, e.g., one variation, two variations, three variations, four variations, or five or more variations in its sequence, as long as the conservative variant

25

30

100

W.

10

15

20

25

30

functionally interacts with a corresponding O-tRNA or O-RS in substantailly the same manner as the non-variant form.

[0051] Selection or screening agent: As used herein, the term "selection or screening agent" refers to an agent that, when present, allows for selection/screening of
5 certain components from a population. For example, a selection or screening agent can be, but is not limited to, e.g., a nutrient, an antibiotic, a wavelength of light, an antibody, an expressed polynucleotide, or the like. The selection agent can be varied, e.g., by concentration, intensity, etc.

Encode: As used herein, the term "encode" refers to any process whereby the information in a polymeric macromolecule or sequence string is used to direct the production of a second molecule or sequence string that is different from the first molecule or sequence string. As used herein, the term is used broadly, and can have a variety of applications. In one aspect, the term "encode" describes the process of semi-conservative DNA replication, where one strand of a double-stranded DNA molecule is used as a template to encode a newly synthesized complementary sister strand by a DNA-dependent DNA polymerase.

[0053] In another aspect, the term "encode" refers to any process whereby the information in one molecule is used to direct the production of a second molecule that has a different chemical nature from the first molecule. For example, a DNA molecule can encode an RNA molecule (e.g., by the process of transcription incorporating a DNA-dependent RNA polymerase enzyme). Also, an RNA molecule can encode a polypeptide, as in the process of translation. When used to describe the process of translation, the term "encode" also extends to the triplet codon that encodes an amino acid. In some aspects, an RNA molecule can encode a DNA molecule, e.g., by the process of reverse transcription incorporating an RNA-dependent DNA polymerase. In another aspect, a DNA molecule can encode a polypeptide, where it is understood that "encode" as used in that case incorporates both the processes of transcription and translation.

BRIEF DESCRIPTION OF THE FIGURES

[0054] Figure 1, Panels A, B and C schematically illustrate leucyl tRNAs and synthetases, and their phylogenetic relationships. Panel A provides a ClustalW analysis of aminoacyl-tRNA synthetases, where Archaeal tRNA synthetases are labeled using a dashed

ŧ

30

line, prokaryotic using a solid line, and eukaryotic sequences using a dotted line. This analysis reveals the halobacterial synthetase to be unusual in its homology to prokaryotic rather than archaeal and eukaryotic synthetases. Panel B provides a ClustalW analysis of Halobacterial tRNAs which all share high homology to other archaeal tRNAs.

- Dendrograms were generated using the program PhyloDraw. Panel C provides a sequence alignment of multiple sequences of the family of archaeal leucyl tRNAs examined as potential orthogonal suppressors. Sequences examined as potential amber suppressors by changing the anticodon (boxed) to CUA are shown in bold as is the consensus sequence. The highly conserved positions G37 and A73 are indicated with underlining.
- 10 [0055] Figure 2 provides a histogram showing the identification of a leucyl orthogonal pair. The suppression efficiency of seven synthetases expressed with 5 orthogonal amber suppressor reporter constructs was measured using a β-lactamase amber suppression assay.
- [0056] Figure 3, Panels A and B provide graphs illustrating aminoacylation in vitro by archaeal leucyl-tRNA synthetases. Panel A illustrates charging of crude total halobacterial tRNA determined by aminoacylation assays with [³H] leucine by AfLRS (■), MjLRS (♠), MtLRS (♠), EcLRS (♠), and no synthetase (□). Panel B illustrates charging of crude total E. coli tRNA.
- [0057] Figure 4, Panels A and B illustrates the optimization of suppressor tRNAs.

 Panel A illustrates regions (shown in boxes) of the halobacterial orthogonal tRNA subjected to mutagenesis in an effort to improve the efficiency or selectivity of TAG and AGGA suppressor tRNAs. Panel B illustrates that active mutant TAG suppressors identified by positive selection conserve A73. Less cross-reactive mutants identified by a double-sieve selection strategy conserve a C3:G70 base pair. The most active and selective suppressor tRNA is shown with double boxes.
 - [0053] Figure 5 illustrates a consensus-derived frameshift suppressor. A consensus sequence was obtained by multiple sequence alignment of all known archaeal leucyl tRNAs, and the anticodon loop is changed to UCUCCUAA. The variations observed for tRNAs identified by selection are shown in boxes. The most active mutations are shown with double boxes.

1.375

5

10

15

20

25

DETAILED DESCRIPTION

[0059] In order to add additional unnatural amino acids to the genetic code *in vivo*, "orthogonal pairs" of an aminoacyl-tRNA synthetase and a tRNA are needed that can function efficiently in the translational machinery. Desired characteristics of the orthogonal pairs include tRNA that decode or recognize only a specific new codon, e.g., a selector codon, that is not decoded by any endogenous tRNA, and aminoacyl-tRNA synthetases that preferentially aminoacylate (or charge) its cognate tRNA with only a specific selected amino acid, e.g., an unnatural amino acid. The O-tRNA is also not typically aminoacylated by endogenous synthetases. For example, in *E. coli*, an orthogonal pair will include an aminoacyl-tRNA synthetase that does not significantly cross-react with any of the endogenous tRNA, which there are 40 in *E. coli*, and an orthogonal tRNA that is not significantly aminoacylated by any of the endogenous synthetases, e.g., of which there are 21 in *E. coli*.

[0060] The O-tRNA is capable of mediating incorporation of a selected amino acid into a protein that is encoded by a polynucleotide, which comprises a selector codon that is recognized by the O-tRNA, e.g., in vivo. The anticodon loop of the O-tRNA recognizes the selector codon on an mRNA and incorporates its amino acid, e.g., a selected amino acid, such as an unnatural amino acid, at this site in the polypeptide. Any of a number of selector codons can be used with the invention. For example, selector codons can include, e.g., nonsense codons, such as, stop codons, e.g., amber, ochre, and opal codons; four or more base codons; rare codons; codons derived from natural or unnatural base pairs and/or the like. See also the section herein entitled "Selector codon."

[0061] By using different selector codons, multiple orthogonal tRNA/synthetase pairs can be developed that allow the simultaneous incorporation of multiple selected amino acids, e.g., unnatural amino acids, using these different selector codons. This invention provides compositions of and methods for identifying and producing additional orthogonal tRNA-aminoacyl-tRNA synthetase pairs, e.g., leucyl O-tRNA/leucyl O-RSs, using any of a number of selector codons, e.g., an amber codon, an opal codon, an extended codon (such as a four-base codon), and the like.

ORTHOGONAL LEUCYL tRNA/ ORTHOGONAL LEUCYL AMINOACYL-tRNA SYNTHETASES AND PAIRS THEREOF

[0062]Such translation systems of the invention generally comprise cells that include an orthogonal leucyl tRNA (leucyl O-tRNA), an orthogonal leucyl aminoacyl tRNA synthetase (leucyl O-RS), and a selected amino acid, e.g., an unnatural amino acid, where - the leucyl O-RS aminoacylates the leucyl O-tRNA with the selected amino acid. An orthogonal pair of the invention is composed of a leucyl O-tRNA, e.g., a suppressor tRNA, a frameshift tRNA, or the like, and an leucyl O-RS. The leucyl-O-tRNA recognize a first selector codon and has at least about a 25% suppression activity in presence of a cognate 10 synthetase in response to a selector codon as compared to a control lacking the cognate synthetase. The leucyl O-tRNA also comprises an anticodon loop comprising a CU(X)_n XXXAA sequence. The cell uses the components to incorporate the selected amino acid into a growing polypeptide chain. For example, a nucleic acid that comprises a polynucleotide that encodes a polypeptide of interest can also be present, where the 15 polynucleotide comprises a selector codon that is recognized by the leucyl O-tRNA. The translation system can also be an in vitro system.

[0063] Translation systems that are suitable for making proteins that include one or more selected amino acids, e.g., an unnatural amino acid, are described in International patent applications WO 2002/086075, entitled "METHODS AND COMPOSITION FOR THE PRODUCTION OF ORTHOGANOL tRNA-AMINOACYLtRNA SYNTHETASE PAIRS" and WO 2002/085923, entitled "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS." In addition, see International Application Number PCT/US2004/011786, filed April 16, 2004. Each of these applications is incorporated herein by reference in its entirety. These translation systems can be adapted to the present invention by substituting the leucyl-O-RS and leucyl-O-tRNA provided herein.

20

25

30

In certain embodiments, a cell of teh invention, e.g., an *E. coli* cell, includes such a translation system of the invention. For example, an *E. coli* cell of the invention can include an orthogonal leucyl-tRNA (leucyl-O-tRNA), where the leucyl-O-tRNA comprises at least about a 25% suppression activity in presence of a cognate synthetase in response to a selector codon as compared to a control lacking the cognate synthetase; an orthogonal leucyl aminoacyl-tRNA synthetase (leucyl-O-RS); a selected amino acid; and, a nucleic

45.54

acid that comprises a polynucleotide that encodes a polypeptide of interest, where the polynucleotide comprises a selector codon that is recognized by the leucyl O-tRNA.

[0065] The invention also features multiple O-tRNA/O-RS pairs in a cell, which allows incorporation of more than one selected amino acid. In certain embodiments, the cell can further include an additional different O-tRNA/O-RS pair and a second selected amino acid, where the O-tRNA recognizes a second selector codon and the O-RS preferentially aminoacylates the O-tRNA with the second selected amino acid. For example, a cell can further comprise, e.g., an amber suppressor tRNA-aminoacyl tRNA synthetase pair derived from the tyrosyl-tRNA synthetase of Methanococcus jannaschii.

10 [0066] The leucyl O-tRNA and/or the leucyl O-RS can be naturally occurring or can be derived by mutation of a naturally occurring tRNA and/or RS, e.g., which generates libraries of tRNAs and/or libraries of RSs, from a variety of organisms. For example, one strategy of producing an orthogonal leucyl tRNA/leucyl aminoacyl-tRNA synthetase pair involves importing a heterologous tRNA/synthetase pair from, e.g., a source other than the host cell, or multiple sources, into the host cell. The properties of the heterologous synthetase candidate include, e.g., that it does not charge any host cell tRNA, and the properties of the heterologous tRNA candidate include, e.g., that it is not aminoacylated by any host cell synthetase. In addition, the heterologous tRNA is orthogonal to all host cell synthetases.

20 [0067] A second strategy for generating an orthogonal pair involves generating mutant libraries from which to screen and/or select a leucyl O-tRNA or leucyl O-RS. These strategies can also be combined.

[0068] In various embodiments, the leucyl O-tRNA and leucyl O-RS are derived from at least one organism. In another embodiment, the leucyl O-tRNA is derived from a naturally occurring or mutated naturally occurring tRNA from a first organism and the leucyl O-RS is derived from naturally occurring or mutated naturally occurring RS from a second organism. In one embodiment, the first and second organism is different. For example, an orthogonal pair of the invention includes a leucyl-tRNA synthetase derived from *Methanobacterium thermoautotrophicum*, and a leucyl tRNA derived from an archael tRNA (e.g., from *Halobacterium sp. NRC-1*). Alternatively, the first and second organism

25

30

are the same. See the section entitled "Sources and Hosts" herein for additional information.

[0069] In certain embodiments of the invention, a leucyl O-tRNA of the invention comprises or is encoded or transcribed by or from a polynucleotide sequence as set forth in any one of SEQ ID NO.: 3, 6, 7 or 12, or a complementary polynucleotide sequence thereof. In certain embodiments, a leucyl O-RS comprises an amino acid sequence as set forth in any one of SEQ ID NO.: 15 or 16, or a conservative variation thereof. The leucyl O-RS, or a portion thereof, can also be encoded or transcribed by or from a polynucleotide sequence as set forth in any one of SEQ ID NO.: 13 or 14, or a complementary polynucleotide sequence thereof. See also, the section entitled "Nucleic Acid and Polypeptide Sequence and Variants," herein.

Orthogonal tRNA (O-tRNA)

10

15

20

25

30

[0070] An orthogonal leucyl tRNA (leucyl O-tRNA) mediates incorporation of a selected amino acid into a protein that is encoded by a polynucleotide that comprises a selector codon that is recognized by the leucyl O-tRNA, e.g., in vivo. A leucyl O-tRNA of the invention comprises an anticodon loop comprising a CU(X)_n XXXAA sequence.

[0071] The $CU(X)_n$ XXXAA sequence is found in the anticodon loop, where X refers to any nucleotide, and $(X)_n$ is optionally present. The n refers to a number of bases the anticodon loop is extended, based on the desired selector codon, e.g., a stop codon (n=0), an extended codon, such as a four- (n=1), five- (n=2), six- (n=3) base pair, etc.

[0072] In one aspect of the invention, the CU(X)_n XXXAA sequence comprises CUCUAAA sequence (n=0), typically when the selector codon is an amber codon. In addition, the leucyl O-tRNA can include a stem region comprising matched base pairs and a conserved discriminator base (position 73). See, e.g., Figure 4, Panel E. This position is indicated in e.g., Figure 4, Panel A. The leucyl O-tRNA also optionally includes a C:G base pair at position 3:70.

In one example, the CU(X)_n XXXAA sequence comprises a CUUCCUAA sequence, typically when the selector codon is a four-base codon. See, e.g., Figure 5. The leucyl O-tRNA can also include a first pair selected from T28:A42, G28:C42 and/or C28:G42, and a second pair selected from G:49:C65 or C49:G65, where the numbering corresponds to that indicated in Figure 4, Panel A. In one embodiment, C28:G42 is the

first pair and C49:G65 is the second pair. When the selector codon is an opal codon, the CU(X)_n XXXAA sequence can comprises a CUUCAAA sequence.

[0074] A leucyl O-tRNA of the invention comprises at least about a 25% suppression activity in presence of a cognate synthetase in response to a selector codon, as 5 _ compared to a control lacking the cognate synthetase. Suppression activity can be determined by any of a number of assays known in the art. For example, a β -galactosidase reporter assay can be used. A derivative of a plasmid that expresses lacZ gene under the control of promoter is used, e.g., where the Leu-25 of the peptide VVLORRDWEN of lacZ is replaced by a selector codon, e.g., TAG, TGA, AGGA, etc. codons, or sense codons (as a 10 control) for tyrosine, serine, leucine, etc. The derivatived lacZ plasmid is introduced into cells from an appropriate organism (e.g., an organism where the orthogonal components can be used) along with plasmid comprising a O-tRNA of the invention. A cognate synthetase can also be introduced (either as a polypeptide or a polynucleotide that encodes the cognate synthetase when expressed). The cells are grown in media to a desired density, e.g., to an 15 OD₆₀₀ of about 0.5., and β -galactosidase assays are performed, e.g., using the BetaFluorTM β-Galactosidase Assay Kit (Novagen). Percent suppression is calculated as the percentage of activity for a sample relative to a comparable control, e.g., the value observed from the derivatived lacZ construct, where the construct has a corresponding sense codon at desired position rather than a selector codon.

20 [0075]Examples of leucyl O-tRNAs of the invention are transcribed from any one of SEQ ID NO.: 1-7 and/or 12. See, Table 3 and Example 2, herein, for sequences of exemplary O-RS and O-tRNA molecules. In the tRNA molecule, Thymine (T) is replace with Uracil (U); the tRNAs have the same sequence, except for the usual substitution of U's for T's. One of skill will appreciate that the RNA and DNA versions of a tRNA are often 25 referred to simply by reference to the DNA sequence that corersponds to the RNA form of the tRNA. Any time a DNA form of a tRNA is given, one of skill will easily be able to derive the RNA (or vice versa) by strandard transcription (or reverse transcription). In addition, additional modifications to the bases can be present. The invention also includes conservative variations of leucyl O-tRNA. For example, conservative variations of leucyl 30 O-tRNA include those molecules that function like the leucyl O-tRNA of any one of SEQ ID NO.: 1-7 and 12 and maintain the tRNA L-shaped structure, but do not have the same

sequence (and are other than wild type leucyl tRNA molecules). See also, the section herein entitled "Nucleic acids and Polypeptides Sequence and Variants."

[0076] The composition comprising a leucyl O-tRNA can further include an orthogonal leucyl aminoacyl-tRNA synthetase (leucyl O-RS), where the leucyl O-RS
5 preferentially aminoacylates the leucyl O-tRNA with a selected amino acid (e.g., an unnatural amino acid). In certain embodiments, a composition that includes a leucyl O-tRNA can further include a translation system (e.g., in vitro or in vivo). A nucleic acid that comprises a polynucleotide that encodes a polypeptide of interest, where the polynucleotide comprises a selector codon that is recognized by the leucyl O-tRNA, or a combination of one or more of these can also be present in the cell. See also, the section herein entitled "Orthogonal aminoacyl-tRNA synthetases."

[0077] Methods of producing an orthogonal tRNA (O-tRNA), e.g., a leucyl O-tRNA, are also a feature of the invention. An O-tRNA, e.g., a leucyl O-tRNA, produced by the method is also a feature of the invention. For example, a method includes mutating an anticodon loop of members of a pool of tRNAs (e.g., a pool of leucyl tRNAs) to allow recognition of a selector codon (e.g., an amber codon, an opal codon, a four base codon, etc.), thereby providing a plurality of potential O-tRNAs; and analyzing secondary structure of a member of the plurality potential O-tRNA to identify non-canonical base pairs in the secondary structure, and optionally mutating the non-canonical base pairs (e.g., the non-canonical base pairs are mutated to canonical base pairs). The non-canonical base pairs can be located in stem region of the secondary structure. Typically, a leucyl O-tRNA possesses an improvement of orthogonality for a desired organism compared to the starting material, e.g., the plurality of tRNA sequences, while preserving its affinity towards a desired RS.

15

20

25

30

[0078] The methods optionally include analyzing the homology of sequences of tRNAs and/or aminoacyl-tRNA synthetases to determine potential candidates for an O-tRNA, O-RS and/or pairs thereof, that appear to be orthogonal for a specific organism. Computer programs known in the art and described herein can be used for the analysis. In one example, to choose potential orthogonal translational components for use in *E. coli*, a prokaryotic organism, a synthetase and/or a tRNA is chosen that does not display unusual homology to prokaryotic organisms.

10.55

5

10

20

25

[0079] The pool of tRNAs can also be produced by a consensus strategy. For example, the pool of tRNAs is produced by aligning a plurality of tRNA sequences (see e.g., Figure 1, Panel C); determining a consensus sequence (see e.g., Figure 1, Panel C); and generating a library of tRNAs using at least a portion, most of, or the entire consensus sequence. For example, a consensus sequence can be compiled with a computer program, e.g., the GCG program pileup. Optionally, degenerate positions determined by the program are changed to the most frequent base at those positions. A library is synthesized by techniques known in the art using the consensus sequence. For example, overlap extension of oligonucleotides in which each site of the tRNA gene can be synthesized as a doped mixture of 90% the consensus sequence and 10% a mixture of the other 3 bases can be used to provide the library based on the consensus sequence. Other mixtures can also be used, e.g., 75% the consensus sequence and 25% a mixture of the other 3 bases, 80% the consensus sequence and 20% a mixture of the other 3 bases, 95% the consensus sequence and 5% a mixture of the other 3 bases, etc.

15 [0080] The library of mutant tRNAs can be generated using various mutagenesis techniques known in the art. For example, the mutant tRNAs can be generated by sitespecific mutations, random point mutations, homologous recombination, DNA shuffling or other recursive mutagenesis methods, chimeric construction or any combination thereof.

[0081]Additional mutations can be introduced at a specific position(s), e.g., at a nonconservative position(s), or at a conservative position, at a randomized position(s), or a combination of both in a desired loop or region of a tRNA, e.g., an anticodon loop, the acceptor stem, D arm or loop, variable loop, T\u00fcC arm or loop, other regions of the tRNA molecule, or a combination thereof. Typically, mutations in a leucyl tRNA include introducing a CU(X)_n XXXAA sequence into the anticodon loop, where X refers to any nucleotide, and $(X)_n$ is optionally present. The n refers to number of bases the anticodon loop needs to be extended based on the selector codon, e.g., an extended codon, such as a four-, five-, six- base pair, etc. In one embodiment, mutations include matched base pairs in the stem region. In one embodiment, mutations include a first pair selected from T28:A42, G28:C42; C28:G42, etc. and a second pair selected from G49:C65 or C49:G65. The 30 numbering refers to the positions on a tRNA molecule, e.g., see Figure 4, Panel A. The

method can further include adding an additional sequence (CCA) to 3' terminus of the O-

tRNA and/or measuring suppression activity.

1

5

10

15

20

25

30

[0082] Typically, an O-tRNA is obtained by subjecting to negative selection a population of cells of a first species, where the cells comprise a member of the plurality of potential O-tRNAs. The negative selection eliminates cells that comprise a member of the plurality of potential O-tRNAs that is aminoacylated by an aminoacyl-tRNA synthetase (RS) that is endogenous to the cells. This provides a pool of tRNAs that are orthogonal to the cell of the first species.

[0083] In certain embodiment in the negative selection, a selector codon(s) is introduced into polynucleotide that encodes a negative selection marker, e.g., an enzyme that confers antibiotic resistance, e.g., β -lactamase, an enzyme that confers a detectable product, e.g., β -galactosidase, chloramphenicol acetyltransferase (CAT), e.g., a toxic product, such as barnase, at a nonessential position, etc. Screening/selection can be done by growing the population of cells in the presence of a selective agent (e.g., an antibiotic, such as ampicillin). In one embodiment, the concentration of the selection agent is varied.

[0084] For example, to measure the activity of suppressor leucyl tRNAs, a selection system is used that is based on the *in vivo* suppression of selector codon, e.g., nonsense or frameshift mutations introduced into a polynucleotide that encodes a negative selection marker, e.g., a gene for β-lactamase (*bla*). For example, polynucleotide variants, e.g., *bla* variants, with, e.g., TAG, AGGA, and TGA, at a certain position (e.g., A184), are constructed. Cells, e.g., bacteria, are transformed with these polynucleotides. In the case of an orthogonal leucyl tRNA, which cannot be efficiently charged by endogenous *E. coli* synthetases, antibiotic resistance, e.g., ampicillin resistance, should be about or less than that for a bacteria transformed with no plasmid. If the leucyl tRNA is not orthogonal, or if a heterologous synthetase capable of charging the tRNA is co-expressed in the system, a higher level of antibiotic, e.g., ampicillin, resistance is be observed. Cells, e.g., bacteria, are chosen that are unable to grow on LB agar plates with antibiotic concentrations about equal to cells transformed with no plasmids.

[0085] In the case of a toxic product (e.g., ribonuclease or barnase), when a member of the plurality of potential leucyl tRNAs is aminoacylated by endogenous host, e.g., *Escherichia coli* synthetases (i.e., it is not orthogonal to the host, e.g., *Escherichia coli* synthetases), the selector codon is suppressed and the toxic polynucleotide product

 $A_{ij} H_{ij}$

20

25

30

produced leads to cell death. Cells harboring orthogonal leucyl tRNAs or non-functional tRNAs survive.

[0086] In one embodiment, the pool of tRNAs that are orthogonal to a desired organism are then subjected to a positive selection in which a selector codon is placed in a 5 _ positive selection marker, e.g., encoded by a drug resistance gene, such a β-lactamase gene. The positive selection is performed on cell comprising a polynucleotide encoding or comprising a member of the pool of tRNAs, a polynucleotide encoding a positive selection marker, and a polynucleotide encoding a cognate RS. These polynucleotides are expressed in the cell and the cell is grown in the presence of a selection agent, e.g., ampicillin. Leucyl 10 tRNAs are then selected for their ability to be aminoacylated by the coexpressed cognate synthetase and to insert an amino acid in response to this selector codon. Typically, these cells show an enhancement in suppression efficiency compared to cells harboring nonfunctional tRNA(s), or tRNAs that cannot efficiently be recognized by the synthetase of interest. The cell harboring the non-functional tRNAs or tRNAs that are not efficiently 15 recognized by the synthetase of interest, are sensitive to the antibiotic. Therefore, leucyl tRNAs that: (i) are not substrates for endogenous host, e.g., Escherichia coli, synthetases; (ii) can be aminoacylated by the synthetase of interest; and (iii) are functional in translation, survive both selections.

[0087] The stringency of the selection, e.g., the positive selection, the negative selection or both the positive and negative selection, in the above described-methods, optionally include varying the selection stringency. For example, because barnase is an extremely toxic protein, the stringency of the negative selection can be controlled by introducing different numbers of selector codons into the barnase gene and/or by using an inducible promoter. In another example, the concentration of the selection or screening agent is varied (e.g., ampicillin concentration). In one aspect of the invention, the stringency is varied because the desired activity can be low during early rounds. Thus, less stringent selection criteria are applied in early rounds and more stringent criteria are applied in later rounds of selection. In certain embodiments, the negative selection, the positive selection or both the negative and positive selection, can be repeated multiple times. Multiple different negative selection markers, positive selection markers or both negative and positive selection markers or both negative and negative selection markers can be the same.

[0088] Other types of selections/screening can be used in the invention for producing orthogonal translational components, e.g., a leucyl O-tRNA, a leucyl O-RS, and a leucyl O-tRNA/O-RS pair. For example, the negative selection marker, the positive selection marker or both the positive and negative selection markers can include a marker that fluoresces or catalyzes a luminescent reaction in the presence of a suitable reactant. In another embodiment, a product of the marker is detected by fluorescence-activated cell sorting (FACS) or by luminescence. Optionally, the marker includes an affinity based screening marker. See, Francisco, J. A., et al., (1993) Production and fluorescence-activated cell sorting of Escherichia coli expressing a functional antibody fragment on the external surface. Proc Natl Acad Sci U S A. 90:10444-8.

[0089] Additional general methods for producing a recombinant orthogonal tRNA can be found, e.g., in International patent applications WO 2002/086075, entitled "Methods and compositions for the production of orthogonal tRNA-aminoacyltRNA synthetase pairs;" and, International Application Number PCT/2004/011786, filed April 16, 2004, entitled "EXPANDING THE EUKARYOTIC GENETIC CODE." See also, Forster et al., (2003) Programming peptidomimetic synthetases by translating genetic codes designed de novo PNAS 100(11):6353-6357; and, Feng et al., (2003), Expanding tRNA recognition of a tRNA synthetase by a single amino acid change, PNAS 100(10): 5676-5681. These are applied to the present invention, e.g., using the substrates (e.g., leucyl-O-tRNAs or O-RSs) in such available selection methods.

Orthogonal aminoacyl-tRNA synthetase (O-RS)

10

15

20

25

30

[0090] A leucyl O-RS of the invention preferentially aminoacylates a leucyl O-tRNA with a selected amino acid in vitro or in vivo. A leucyl O-RS of the invention can be provided to the translation system, e.g., a cell, by a polypeptide that includes a leucyl O-RS and/or by a polynucleotide that encodes a leucyl O-RS or a portion thereof. For example, a leucyl O-RS, or a portion thereof, is encoded by a polynucleotide sequence as set forth in any one of SEQ ID NO.: 13-14, or a complementary polynucleotide sequence thereof. In another example, a leucyl O-RS comprises an amino acid sequence as set forth in any one of SEQ ID NO.: 15-16, or a conservative variation thereof. See, e.g., Table 3 and Example 2 herein for sequences of exemplary leucyl O-RS molecules.

[0091] Methods for identifying an orthogonal aminoacyl-tRNA synthetase (O-RS), e.g., a leucyl O-RS, for use with an O-tRNA, e.g., a leucyl O-tRNA, are also a feature of the

invention. For example, a method includes subjecting to positive selection a population of cells of a first species, where the cells individually comprise: 1) a member of a plurality of aminoacyl-tRNA synthetases (RSs), where the plurality of RSs comprise mutant RSs. RSs derived from a species other than the first species or both mutant RSs and RSs derived from a species other than the first species; 2) the orthogonal tRNA (O-tRNA) from a second species; and 3) a polynucleotide that encodes a positive selection marker and comprises at least one selector codon. Cells are selected or screened for those that show an enhancement in suppression efficiency compared to cells lacking or with a reduced amount of the member of the plurality of RSs. Cells having an enhancement in suppression efficiency comprise an active RS that aminoacylates the O-tRNA. A level of aminoacylation (in vitro or in vivo) by the active RS of a first set of tRNAs from the first species is compared to the level of aminoacylation (in vitro or in vivo) by the active RS of a second set of tRNAs from the second species. The level of aminoacylation can be determined by a detectable substance (e.g., a labeled amino acid or unnatural amino acid). The active RS that more efficiently aminoacylates the second set of tRNAs compared to the first set of tRNAs is selected, thereby providing an efficient (optimized) orthogonal aminoacyl-tRNA synthetase for use with the O-tRNA. An O-RS, e.g., a leucyl O-RS, identified by the method, is also a feature of the invention.

10

15

30

[0092] Any of a number of assays can be used to determine aminoacylation. These assays can be performed *in vitro* or *in vivo*. For example, in vitro aminoacylation assays are described in, e.g., Hoben, P., and Soll, D. (1985) Methods Enzymol. 113:55-59. Aminoacylation can also be determined by using a reporter along with orthogonal translation components and detecting the reporter in a cell expressing a polynucleotide comprising at least one selector codon that encodes a protein. See also, WO 2002/085923, entitled "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS;" and, USSN 60/479,931 entitled "EXPANDING THE EUKARYOTIC GENETIC CODE."

[0093] Identified leucyl O-RS can be further manipulated to alter the substrate specificity of the synthetase, so that only a desired unnatural amino acid, but not any of the common 20 amino acids are charged to the leucyl O-tRNA. Methods to generate an orthogonal leucyl aminoacyl tRNA synthetase with a substrate specificity for an unnatural amino acid include mutating the synthetase, e.g., at the active site in the synthetase, at the editing mechanism site in the synthetase, at different sites by combining different domains

of synthetases, or the like, and applying a selection process. A strategy is used, which is based on the combination of a positive selection followed by a negative selection. In the positive selection, suppression of the selector codon introduced at a nonessential position(s) of a positive marker allows cells to survive under positive selection pressure. In the presence of both natural and unnatural amino acids, survivors thus encode active synthetases charging the orthogonal suppressor tRNA with either a natural or unnatural amino acid. In the negative selection, suppression of a selector codon introduced at a nonessential position(s) of a negative marker removes synthetases with natural amino acid specificities. Survivors of the negative and positive selection encode synthetases that aminoacylate (charge) the orthogonal suppressor tRNA with unnatural amino acids only. These synthetases can then be subjected to further mutagenesis, e.g., DNA shuffling or other recursive mutagenesis methods.

10

15

20

mutagenesis techniques known in the art. For example, the mutant RSs can be generated by site-specific mutations, random point mutations, homologous recombination, DNA shuffling or other recursive mutagenesis methods, chimeric construction or any combination thereof. For example, a library of mutant leucyl RSs can be produced from two or more other, e.g., smaller, less diverse "sub-libraries." Chimeric libraries of RSs are also included in the invention. It should be noted that libraries of tRNA synthetases from various organism (e.g., microorganisms such as eubacteria or archaebacteria) such as libraries that comprise natural diversity (see, e.g., U.S. Patent No. 6,238,884 to Short et al; U.S. Patent No. 5,756,316 to Schallenberger et al; U.S. Patent No. 5,783,431 to Petersen et al; U.S. Patent No. 5,824,485 to Thompson et al; U.S. Patent No. 5,958,672 to Short et al), are optionally constructed and screened for orthogonal pairs.

25 [0095] Once the synthetases are subject to the positive and negative selection/screening strategy, these synthetases can then be subjected to further mutagenesis. For example, a nucleic acid that encodes the leucyl O-RS can be isolated; a set of polynucleotides that encode mutated leucyl O-RSs (e.g., by random mutagenesis, site-specific mutagenesis, recombination or any combination thereof) can be generated from the nucleic acid; and, these individual steps or a combination of these steps can be repeated until a mutated leucyl O-RS is obtained that preferentially aminoacylates the leucyl O-

tRNA with the unnatural amino acid. In one aspect of the invention, the steps are performed multiple times, e.g., at least two times.

[0096] Additional levels of selection/screening stringency can also be used in the methods of the invention, for producing leucyl O-tRNA, leucyl O-RS, or pairs thereof. The selection or screening stringency can be varied on one or both steps of the method to produce an O-RS. This could include, e.g., varying the amount of selection/screening agent that is used, etc. Additional rounds of positive and/or negative selections can also be performed. Selecting or screening can also comprise one or more positive or negative selection or screening that includes, e.g., a change in amino acid permeability, a change in translation efficiency, a change in translational fidelity, etc. Typically, the one or more change is based upon a mutation in one or more gene in an organism in which an orthogonal tRNA-tRNA synthetase pair is used to produce protein.

[0097] Additional general details for producing O-RS, and altering the substrate specificity of the synthetase can be found in WO 2002/086075 entitled "Methods and compositions for the production of orthogonal tRNA-aminoacyltRNA synthetase pairs;" and International Application Number PCT/US2004/011786, filed April 16, 2004.

SOURCE AND HOST ORGANISMS

10

15

[0098]The translational components of the invention can be derived from noneukaryotic organisms. For example, the orthogonal O-tRNA can be derived from a non-20 eukaryotic organism, e.g., an archaebacterium, such as Methanococcus jannaschii, Methanobacterium thermoautotrophicum, Halobacterium such as Haloferax volcanii and Halobacterium species NRC-1, Archaeoglobus fulgidus, Pyrococcus furiosus, Pyrococcus horikoshii, Aeuropyrum pernix, or the like, or a eubacterium, such as Escherichia coli, Thermus thermophilus, Bacillus stearothermphilus, or the like, while the orthogonal O-RS 25 can be derived from a non-eukaryotic organism, e.g., Methanobacterium thermoautotrophicum, Halobacterium such as Haloferax volcanii and Halobacterium species NRC-1, Archaeoglobus fulgidus, Pyrococcus furiosus, Pyrococcus horikoshii, Aeuropyrum pernix, or the like, or a eubacterium, such as Escherichia coli, Thermus thermophilus, Bacillus stearothermphilus, or the like. In one embodiment, eukaryotic 30 sources can also be used, e.g., plants, algae, protists, fungi, yeasts, animals (e.g., mammals, insects, arthropods, etc.), or the like.

[0099] The individual components of a leucyl O-tRNA/O-RS pair can be derived from the same organism or different organisms. In one embodiment, the leucyl O-tRNA/O-RS pair is from the same organism. Alternatively, the leucyl O-tRNA and the leucyl O-RS of the leucyl O-tRNA/O-RS pair are from different organisms. For example, the leucyl O-tRNA can be derived from, e.g., a *Halobacterium sp NRC-1*, and the leucyl O-RS can be derived from, e.g., a *Methanobacterium thermoautrophicum*.

[0100]The leucyl O-tRNA, leucyl O-RS or leucyl O-tRNA/O-RS pair can be selected or screened in vivo or in vitro and/or used in a cell, e.g., a non-eukaryotic cells (such as E. coli cell), or a eukaryotic cell, to produce a polypeptide with a selected amino 10 acid (e.g., an unnatural amino acid). A non-eukaryotic cell can be from a variety of sources, e.g., Methanobacterium thermoautotrophicum, Halobacterium such as Haloferax volcanii and Halobacterium species NRC-1, Archaeoglobus fulgidus, Pyrococcus furiosus, Pyrococcus horikoshii, Aeuropyrum pernix, or the like, or a eubacterium, such as Escherichia coli, Thermus thermophilus, Bacillus stearothermphilus, or the like. A 15 eukaryotic cell can be from any of a variety of sources, e.g., a plant (e.g., complex plant such as monocots, or dicots), an algae, a protist, a fungus, a yeast (e.g., Saccharomyces cerevisiae), an animal (e.g., a mammal, an insect, an arthropod, etc.), or the like. Compositions of cells with translational components of the invention are also a feature of the invention.

20 [0101] See also, International Application Number PCT/US2004/011786, filed April 16, 2004, entitled "Expanding the Eukaryotic Genetic Code" for screening O-tRNA and/or O-RS in one species for use in another species.

SELECTOR CODONS

[0102] Selector codons of the invention expand the genetic codon framework of protein biosynthetic machinery. For example, a selector codon includes, e.g., a unique three base codon, a nonsense codon, such as a stop codon, e.g., an amber codon (UAG), or an opal codon (UGA), an unnatural codon, at least a four base codon, a rare codon, or the like. A number of selector codons can be introduced into a desired gene, e.g., one or more, two or more, more than three, etc.

In one embodiment, the methods involve the use of a selector codon that is a stop codon for the incorporation of a selected amino acid, e.g., an unnatural amino acids, in

mg -

20

25

30

vivo in a cell. For example, a leucyl O-tRNA is produced that recognizes the stop codon and is aminoacylated by a leucyl O-RS with a selected amino acid. This leucyl O-tRNA is not recognized by the naturally occurring host's aminoacyl-tRNA synthetases. Conventional site-directed mutagenesis can be used to introduce the stop codon at the site of interest in a polypeptide of interest. See, e.g., Sayers, J.R., et al. (1988), 5',3' Exonuclease 5 in phosphorothioate-based oligonucleotide-directed mutagenesis. Nucleic Acids Res, 791-802. When the leucyl O-RS, leucyl O-tRNA and the nucleic acid that encodes a polypeptide of interest are combined, e.g., in vivo, the selected amino acid is incorporated in response to the stop codon to give a polypeptide containing the selected amino acid, e.g., an unnatural amino acid, at the specified position. In one embodiment of the invention, a stop 10 codon used as a selector codon is an amber codon, UAG, and/or an opal codon, UGA. For example, see SEQ ID NO: 3 for an example of a leucyl O-tRNA that recognizes an amber codon, and see SEQ ID NO: 7 for an example of a leucyl O-tRNA that recognizes an opal codon. A genetic code in which UAG and UGA are both used as a selector codon can encode 22 amino acids while preserving the ochre nonsense codon, UAA, which is the most 15 abundant termination signal.

[0104] The incorporation of selected amino acids, e.g., unnatural amino acids, in vivo, can be done without significant perturbation of the host cell. For example, in non-eukaryotic cells, such as Escherichia coli, because the suppression efficiency for the UAG codon depends upon the competition between the O-tRNA, e.g., the amber suppressor tRNA, and the release factor 1 (RF1) (which binds to the UAG codon and initiates release of the growing peptide from the ribosome), the suppression efficiency can be modulated by, e.g., either increasing the expression level of O-tRNA, e.g., the suppressor tRNA, or using an RF1 deficient strain. In eukaryotic cells, because the suppression efficiency for the UAG codon depends upon the competition between the O-tRNA, e.g., the amber suppressor tRNA, and a eukaryotic release factor (e.g., eRF) (which binds to a stop codon and initiates release of the growing peptide from the ribosome), the suppression efficiency can be modulated by, e.g., increasing the expression level of O-tRNA, e.g., the suppressor tRNA.

[0105] Unnatural amino acids can also be encoded with rare codons. For example, when the arginine concentration in an in vitro protein synthesis reaction is reduced, the rare arginine codon, AGG, has proven to be efficient for insertion of Ala by a synthetic tRNA acylated with alanine. See, e.g., Ma et al., Biochemistry, 32:7939 (1993). In this case, the

synthetic tRNA competes with the naturally occurring tRNAArg, which exists as a minor species in *Escherichia coli*. Some organisms do not use all triplet codons. An unassigned codon AGA in *Micrococcus luteus* has been utilized for insertion of amino acids in an in vitro transcription/translation extract. *See*, e.g., Kowal and Oliver, Nucl. Acid. Res., 25:4685 (1997). Components of the present invention can be generated to use these rare codons in vivo.

5

[0106]Selector codons can also comprise extended codons, e.g., four or more base codons, such as, four, five, six or more base codons. Examples of four base codons include, e.g., AGGA, CUAG, UAGA, CCCU, and the like. Examples of five base codons include, e.g., AGGAC, CCCCU, CCCUC, CUAGA, CUACU, UAGGC, and the like. Methods of 10 the invention include using extended codons based on frameshift suppression. Four or more base codons can insert, e.g., one or multiple selected amino acids, e.g., unnatural amino acids, into the same protein. For example, in the presence of mutated leucyl O-tRNAs, e.g., a special frameshift suppressor tRNAs, with anticodon loops, e.g., with a CU(X) n XXXAA 15 sequence (where n=1), the four or more base codon is read as single amino acid. For example, see SEQ ID NOs.: 6 and 12 for leucyl O-tRNAs that recognize a four base codon. In other embodiments, the anticodon loops can decode, e.g., at least a four-base codon, at least a five-base codon, or at least a six-base codon or more. Since there are 256 possible four-base codons, multiple unnatural amino acids can be encoded in the same cell using a 20 four or more base codon. See also, Anderson et al., (2002) Exploring the Limits of Codon and Anticodon Size, Chemistry and Biology, 9:237-244; Magliery, (2001) Expanding the Genetic Code: Selection of Efficient Suppressors of Four-base Codons and Identification of "Shifty" Four-base Codons with a Library Approach in Escherichia coli, J. Mol. Biol. 307: 755-769.

25 [0107] For example, four-base codons have been used to incorporate unnatural amino acids into proteins using in vitro biosynthetic methods. See, e.g., Ma et al., (1993)

Biochemistry, 32:7939; and Hohsaka et al., (1999) J. Am. Chem. Soc., 121:34. CGGG and AGGU were used to simultaneously incorporate 2-naphthylalanine and an NBD derivative of lysine into streptavidin in vitro with two chemically acylated frameshift suppressor tRNAs. See, e.g., Hohsaka et al., (1999) J. Am. Chem. Soc., 121:12194. In an in vivo study, Moore et al. examined the ability of tRNALeu derivatives with NCUA anticodons to suppress UAGN codons (N can be U, A, G, or C), and found that the quadruplet UAGA can

be decoded by a tRNALeu with a UCUA anticodon with an efficiency of 13 to 26% with little decoding in the 0 or -1 frame. See, Moore et al., (2000) J. Mol. Biol., 298:195. In one embodiment, extended codons based on rare codons or nonsense codons can be used in invention, which can reduce missense readthrough and frameshift suppression at other unwanted sites.

5

- [0108] For a given system, a selector codon can also include one of the natural three base codons, where the endogenous system does not use (or rarely uses) the natural base codon. For example, this includes a system that is lacking a tRNA that recognizes the natural three base codon, and/or a system where the three base codon is a rare codon.
- 10 [0109] Selector codons optionally include unnatural base pairs. These unnatural base pairs further expand the existing genetic alphabet. One extra base pair increases the number of triplet codons from 64 to 125. Properties of third base pairs include stable and selective base pairing, efficient enzymatic incorporation into DNA with high fidelity by a polymerase, and the efficient continued primer extension after synthesis of the nascent unnatural base pair. Descriptions of unnatural base pairs which can be adapted for methods and compositions include, e.g., Hirao, et al., (2002) An unnatural base pair for incorporating amino acid analogues into protein, Nature Biotechnology, 20:177-182. See, also, Wu, Y., et al., (2002) J. Am. Chem. Soc. 124:14626-14630. Other relevant publications are listed herein.
- 20 [0110] For in vivo usage, the unnatural nucleoside is membrane permeable and is phosphorylated to form the corresponding triphosphate. In addition, the increased genetic information is stable and not destroyed by cellular enzymes. Previous efforts by Benner and others took advantage of hydrogen bonding patterns that are different from those in canonical Watson-Crick pairs, the most noteworthy example of which is the iso-C:iso-G 25 pair. See, e.g., Switzer et al., (1989) J. Am. Chem. Soc., 111:8322; and Piccirilli et al., (1990) Nature, 343:33; Kool, (2000) Curr. Opin. Chem. Biol., 4:602. These bases in general mispair to some degree with natural bases and cannot be enzymatically replicated. Kool and co-workers demonstrated that hydrophobic packing interactions between bases can replace hydrogen bonding to drive the formation of base pair. See, Kool, (2000) Curr. 30 Opin. Chem. Biol., 4:602; and Guckian and Kool, (1998) Angew. Chem. Int. Ed. Engl., 36, 2825. In an effort to develop an unnatural base pair satisfying all the above requirements. Schultz, Romesberg and co-workers have systematically synthesized and studied a series of

• • • •

unnatural hydrophobic bases. A PICS:PICS self-pair is found to be more stable than natural base pairs, and can be efficiently incorporated into DNA by Klenow fragment of Escherichia coli DNA polymerase I (KF). See, e.g., McMinn et al., (1999) J. Am. Chem. Soc., 121:11586; and Ogawa et al., (2000) J. Am. Chem. Soc., 122:3274. A 3MN:3MN self-pair can be synthesized by KF with efficiency and selectivity sufficient for biological function. See, e.g., Ogawa et al., (2000) J. Am. Chem. Soc., 122:8803. However, both bases act as a chain terminator for further replication. A mutant DNA polymerase has been recently evolved that can be used to replicate the PICS self pair. In addition, a 7AI self pair can be replicated. See, e.g., Tae et al., (2001) J. Am. Chem. Soc., 123:7439. A novel metallobase pair, Dipic:Py, has also been developed, which forms a stable pair upon binding Cu(II). See, Meggers et al., (2000) J. Am. Chem. Soc., 122:10714. Because extended codons and unnatural codons are intrinsically orthogonal to natural codons, the methods of the invention can take advantage of this property to generate orthogonal tRNAs for them.

[0111] A translational bypassing system can also be used to incorporate a selected amino acid, e.g., an unnatural amino acid, in a desired polypeptide. In a translational bypassing system, a large sequence is inserted into a gene but is not translated into protein. The sequence contains a structure that serves as a cue to induce the ribosome to hop over the sequence and resume translation downstream of the insertion.

SELECTED AND UNNATURAL AMINO ACIDS

15

- 20 [0112] As used herein, a selected amino acid refers to any desired naturally occurring amino acid or unnatural amino acid. A naturally occurring amino acid includes any one of the twenty genetically encoded alpha-amino acids: alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine.

 In one embodiment, the selected amino acid is incorporated into a growing polypeptide chain with high fidelity, e.g., at greater than 75% efficiency for a given selector codon, at greater than about 80% efficiency for a given selector codon, at greater than about 90% efficiency for a given selector codon, or at greater than about 99% or more efficiency for a given selector codon.
- As used herein, an unnatural amino acid refers to any amino acid, modified amino acid, or amino acid analogue other than selenocysteine and/or pyrrolysine and the

·11.

following twenty genetically encoded alpha-amino acids: alanine, arginine, asparagine, aspartic acid, cysteine, glutamine, glutamic acid, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine. The generic structure of an alpha-amino acid is illustrated by Formula I:

I H_2N $C_{Q_2}H$

5

10

15

20

25

[0114] An unnatural amino acid is typically any structure having Formula I wherein the R group is any substituent other than one used in the twenty natural amino acids. See, e.g., *Biochemistry* by L. Stryer, 3rd ed. 1988, Freeman and Company, New York, for structures of the twenty natural amino acids. Note that, the unnatural amino acids of the invention can be naturally occurring compounds other than the twenty alpha-amino acids above.

[0115] Because the unnatural amino acids of the invention typically differ from the natural amino acids in side chain only, the unnatural amino acids form amide bonds with other amino acids, e.g., natural or unnatural, in the same manner in which they are formed in naturally occurring proteins. However, the unnatural amino acids have side chain groups that distinguish them from the natural amino acids.

[0116] Because the unnatural amino acids of the invention typically differ from the natural amino acids in side chain, the unnatural amino acids form amide bonds with other amino acids, e.g., natural or unnatural, in the same manner in which they are formed in naturally occurring proteins. However, the unnatural amino acids have side chain groups that distinguish them from the natural amino acids. For example, R in Formula I optionally comprises an alkyl-, aryl-, acyl-, keto-, azido-, hydroxyl-, hydrazine, cyano-, halo-, hydrazide, alkenyl, alkynl, ether, thiol, seleno-, sulfonyl-, borate, boronate, phospho, phosphono, phosphine, heterocyclic, enone, imine, aldehyde, ester, thioacid, hydroxylamine, amine, and the like, or any combination thereof. In some embodiments, the unnatural amino acids have a photoactivatable cross-linker that is used, e.g., to link a

protein to a solid support. In one embodiment, the unnatural amino acids have a saccharide moiety attached to the amino acid side chain.

[0117] In addition to unnatural amino acids that contain novel side chains, unnatural amino acids also optionally comprise modified backbone structures, e.g., as illustrated by the structures of Formula II and III:

Ш

10

15

20

wherein Z typically comprises OH, NH₂, SH, NH-R', or S-R'; X and Y, which can be the same or different, typically comprise S or O, and R and R', which are optionally the same or different, are typically selected from the same list of constituents for the R group described above for the unnatural amino acids having Formula I as well as hydrogen. For example, unnatural amino acids of the invention optionally comprise substitutions in the amino or carboxyl group as illustrated by Formulas II and III. Unnatural amino acids of this type include, but are not limited to, α -hydroxy acids, α -thioacids α -aminothiocarboxylates, e.g., with side chains corresponding to the common twenty natural amino acids or unnatural side chains. In addition, substitutions at the α -carbon optionally include L, D, or α - α -disubstituted amino acids such as D-glutamate, D-alanine, D-methyl-O-tyrosine, aminobutyric acid, and the like. Other structural alternatives include cyclic amino acids, such as proline analogues as well as 3,4,6,7,8, and 9 membered ring proline analogues, β and γ amino acids such as substituted β -alanine and γ -amino butyric acid.

i

For example, many unnatural amino acids are based on natural amino acids, [0118]such as tyrosine, glutamine, phenylalanine, and the like. Tyrosine analogs include parasubstituted tyrosines, ortho-substituted tyrosines, and meta substituted tyrosines, wherein the substituted tyrosine comprises an acetyl group, a benzoyl group, an amino group, a hydrazine, an hydroxyamine, a thiol group, a carboxy group, an isopropyl group, a methyl group, a C₆ - C₂₀ straight chain or branched hydrocarbon, a saturated or unsaturated hydrocarbon, an O-methyl group, a polyether group, a nitro group, or the like. In addition, multiply substituted aryl rings are also contemplated. Glutamine analogs of the invention include, but are not limited to, α -hydroxy derivatives, γ -substituted derivatives, cyclic derivatives, and amide substituted glutamine derivatives. Example phenylalanine analogs 10 include, but are not limited to, para-substituted phenylalanines, ortho-substituted phenyalanines, and meta-substituted phenylalanines, wherein the substituent comprises a hydroxy group, a methoxy group, a methyl group, an allyl group, an aldehyde or keto group, or the like. Specific examples of unnatural amino acids include, but are not limited to, a pacetyl-L- phenylalanine, a p-propargyl-phenylalanine, O-methyl-L-tyrosine, an L-3-(2-15 naphthyl)alanine, a 3-methyl-phenylalanine, an O-4-allyl-L-tyrosine, a 4-propyl-L-tyrosine, a tri-O-acetyl-GlcNAcβ-serine, an L-Dopa, a fluorinated phenylalanine, an isopropyl-Lphenylalanine, a p-azido-L-phenylalanine, a p-acyl-L-phenylalanine, a p-benzoyl-Lphenylalanine, an L-phosphoserine, a phosphonoserine, a phosphonotyrosine, a p-iodophenylalanine, a p-bromophenylalanine, a p-amino-L-phenylalanine, and an isopropyl-L-20 phenylalanine, and the like. The structures of a variety of unnatural amino acids are provided in, for example, Figures 16, 17, 18, 19, 26, and 29 of WO 2002/085923 entitled "In vivo incorporation of unnatural amino acids."

Chemical Synthesis of Unnatural Amino Acids
Many of the unnatural amino acids provided above are commercially
available, e.g., from Sigma (USA) or Aldrich (Milwaukee, WI, USA). Those that are not
commercially available are optionally synthesized as provided in various publications or
using standard methods known to those of skill in the art. For organic synthesis techniques,
see, e.g., Organic Chemistry by Fessendon and Fessendon, (1982, Second Edition, Willard
Grant Press, Boston Mass.); Advanced Organic Chemistry by March (Third Edition, 1985,
Wiley and Sons, New York); and Advanced Organic Chemistry by Carey and Sundberg
(Third Edition, Parts A and B, 1990, Plenum Press, New York). Additional publications

describing the synthesis of unnatural amino acids include, e.g., WO 2002/085923 entitled "In vivo incorporation of Unnatural Amino Acids;" Matsoukas et al., (1995) J. Med. Chem., 38, 4660-4669; King, F.E. & Kidd, D.A.A. (1949) A New Synthesis of Glutamine and of y-Dipeptides of Glutamic Acid from Phthylated Intermediates. J. Chem. Soc., 3315-3319; 5 Friedman, O.M. & Chatterrii, R. (1959) Synthesis of Derivatives of Glutamine as Model Substrates for Anti-Tumor Agents. J. Am. Chem. Soc. 81, 3750-3752; Craig, J.C. et al. (1988) Absolute Configuration of the Enantiomers of 7-Chloro-4 [[4-(diethylamino)-1methylbutyl]amino]quinoline (Chloroquine). J. Org. Chem. 53, 1167-1170; Azoulay, M., Vilmont, M. & Frappier, F. (1991) Glutamine analogues as Potential Antimalarials,. Eur. J. 10 Med. Chem. 26, 201-5; Koskinen, A.M.P. & Rapoport, H. (1989) Synthesis of 4-Substituted Prolines as Conformationally Constrained Amino Acid Analogues. J. Org. Chem. 54, 1859-1866; Christie, B.D. & Rapoport, H. (1985) Synthesis of Optically Pure Pipecolates from L-Asparagine. Application to the Total Synthesis of (+)-Apovincamine through Amino Acid Decarbonylation and Iminium Ion Cyclization, J. Org. Chem. 1989:1859-1866; Barton et 15 al., (1987) Synthesis of Novel a-Amino-Acids and Derivatives Using Radical Chemistry: Synthesis of L- and D-a-Amino-Adipic Acids, L-a-aminopimelic Acid and Appropriate Unsaturated Derivatives. Tetrahedron Lett. 43:4297-4308; and, Subasinghe et al., (1992) Quisqualic acid analogues: synthesis of beta-heterocyclic 2-aminopropanoic acid derivatives and their activity at a novel quisqualate-sensitized site. J. Med. Chem. 35:4602-20 7. See also, International Application Number PCT/US03/41346, entitled "Protein Arrays," filed on December 22, 2003.

Cellular uptake of unnatural amino acids

25

30

[0120] Unnatural amino acid uptake by a cell is one issue that is typically considered when designing and selecting unnatural amino acids, e.g., for incorporation into a protein. For example, the high charge density of α-amino acids suggests that these compounds are unlikely to be cell permeable. Natural amino acids are taken up into the cell via a collection of protein-based transport systems often displaying varying degrees of amino acid specificity. A rapid screen can be done which assesses which unnatural amino acids, if any, are taken up by cells. *See, e.g.*, the toxicity assays in, e.g., International Application Number PCT/US03/41346, entitled "Protein Arrays," filed on December 22, 2003; and Liu, D.R. & Schultz, P.G. (1999) *Progress toward the evolution of an organism with an expanded genetic code*. PNAS United States 96:4780-4785. Although uptake is

easily analyzed with various assays, an alternative to designing unnatural amino acids that are amenable to cellular uptake pathways is to provide biosynthetic pathways to create amino acids *in vivo*.

Biosynthesis of Unnatural Amino Acids

- 5 **[0121]** Many biosynthetic pathways already exist in cells for the production of amino acids and other compounds. While a biosynthetic method for a particular unnatural amino acid may not exist in nature, e.g., in a cell, the invention provides such methods. For example, biosynthetic pathways for unnatural amino acids are optionally generated in host cell by adding new enzymes or modifying existing host cell pathways. Additional new 10 enzymes are optionally naturally occurring enzymes or artificially evolved enzymes. For example, the biosynthesis of p-aminophenylalanine (as presented in an example in WO 2002/085923, supra) relies on the addition of a combination of known enzymes from other organisms. The genes for these enzymes can be introduced into a cell by transforming the cell with a plasmid comprising the genes. The genes, when expressed in the cell, provide an 15 enzymatic pathway to synthesize the desired compound. Examples of the types of enzymes that are optionally added are provided in the examples below. Additional enzymes sequences are found, e.g., in Genbank. Artificially evolved enzymes are also optionally added into a cell in the same manner. In this manner, the cellular machinery and resources of a cell are manipulated to produce unnatural amino acids.
- 20 [0122] Indeed, any of a variety of methods can be used for producing novel enzymes for use in biosynthetic pathways, or for evolution of existing pathways, for the production of unnatural amino acids, in vitro or in vivo. Many available methods of evolving enzymes and other biosynthetic pathway components can be applied to the present invention to produce unnatural amino acids (or, indeed, to evolve synthetases to have new 25 substrate specificities or other activities of interest). For example, DNA shuffling is optionally used to develop novel enzymes and/or pathways of such enzymes for the production of unnatural amino acids (or production of new synthetases), in vitro or in vivo. See, e.g., Stemmer (1994), Rapid evolution of a protein in vitro by DNA shuffling, Nature 370(4):389-391; and, Stemmer, (1994), DNA shuffling by random fragmentation and 30 reassembly: In vitro recombination for molecular evolution, Proc. Natl. Acad. Sci. USA., 91:10747-10751. A related approach shuffles families of related (e.g., homologous) genes to quickly evolve enzymes with desired characteristics. An example of such "family gene

5

10

15

20

25

shuffling" methods is found in Crameri et al. (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" Nature, 391(6664): 288-291. New enzymes (whether biosynthetic pathway components or synthetases) can also be generated using a DNA recombination procedure known as "incremental truncation for the creation of hybrid enzymes" ("ITCHY"), e.g., as described in Ostermeier et al. (1999) "A combinatorial approach to hybrid enzymes independent of DNA homology" Nature Biotech 17:1205. This approach can also be used to generate a library of enzyme or other pathway variants which can serve as substrates for one or more in vitro or in vivo recombination methods. See, also, Ostermeier et al. (1999) "Combinatorial Protein Engineering by Incremental Truncation," Proc. Natl. Acad. Sci. USA, 96: 3562-67, and Ostermeier et al. (1999), "Incremental Truncation as a Strategy in the Engineering of Novel Biocatalysts." Biological and Medicinal Chemistry, 7: 2139-44. Another approach uses exponential ensemble mutagenesis to produce libraries of enzyme or other pathway variants that are, e.g., selected for an ability to catalyze a biosynthetic reaction relevant to producing an unnatural amino acid (or a new synthetase). In this approach, small groups of residues in a sequence of interest are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. Examples of such procedures, which can be adapted to the present invention to produce new enzymes for the production of unnatural amino acids (or new synthetases) are found in Delegrave & Youvan (1993) Biotechnology Research 11:1548-1552. In yet another approach, random or semi-random mutagenesis using doped or degenerate oligonucleotides for enzyme and/or pathway component engineering can be used, e.g., by using the general mutagenesis methods of e.g., Arkin and Youvan (1992) "Optimizing nucleotide mixtures to encode specific subsets of amino acids for semi-random mutagenesis" <u>Biotechnology</u> 10:297-300; or Reidhaar-Olson et al. (1991) "Random mutagenesis of protein sequences using oligonucleotide cassettes" Methods Enzymol. 208:564-86. Yet another approach, often termed a "non-stochastic" mutagenesis, which uses polynucleotide reassembly and site-saturation mutagenesis can be used to produce enzymes and/or pathway components, which can then be screened for an ability to perform one or more synthetase or biosynthetic pathway function (e.g., for the production of 30 unnatural amino acids in vivo). See, e.g., Short "Non-Stochastic Generation of Genetic Vaccines and Enzymes" WO 00/46344.

[0123] An alternative to such mutational methods involves recombining entire genomes of organisms and selecting resulting progeny for particular pathway functions (often referred to as "whole genome shuffling"). This approach can be applied to the present invention, e.g., by genomic recombination and selection of an organism (e.g., an *E. coli* or other cell) for an ability to produce an unnatural amino acid (or intermediate thereof). For example, methods taught in the following publications can be applied to pathway design for the evolution of existing and/or new pathways in cells to produce unnatural amino acids in vivo: Patnaik et al. (2002) "Genome shuffling of lactobacillus for improved acid tolerance" Nature Biotechnology, 20(7): 707-712; and Zhang et al. (2002) "Genome shuffling leads to rapid phenotypic improvement in bacteria" Nature, February 7, 415(6872): 644-646.

5

10

15

20

25

30

[0124] Other techniques for organism and metabolic pathway engineering, e.g., for the production of desired compounds are also available and can also be applied to the production of unnatural amino acids. Examples of publications teaching useful pathway engineering approaches include: Nakamura and White (2003) "Metabolic engineering for the microbial production of 1,3 propanediol" Curr. Opin. Biotechnol. 14(5):454-9; Berry et al. (2002) "Application of Metabolic Engineering to improve both the production and use of Biotech Indigo" J. Industrial Microbiology and Biotechnology 28:127-133; Banta et al. (2002) "Optimizing an artificial metabolic pathway: Engineering the cofactor specificity of Corynebacterium 2,5-diketo-D-gluconic acid reductase for use in vitamin C biosynthesis" Biochemistry, 41(20), 6226-36; Selivonova et al. (2001) "Rapid Evolution of Novel Traits in Microorganisms" Applied and Environmental Microbiology, 67:3645, and many others.

[0125] Regardless of the method used, typically, the unnatural amino acid produced with an engineered biosynthetic pathway of the invention is produced in a concentration sufficient for efficient protein biosynthesis, e.g., a natural cellular amount, but not to such a degree as to significantly affect the concentration of other cellular amino acids or to exhaust cellular resources. Typical concentrations produced *in vivo* in this manner are about 10 mM to about 0.05 mM. Once a cell is engineered to produce enzymes desired for a specific pathway and an unnatural amino acid is generated, *in vivo* selections are optionally used to further optimize the production of the unnatural amino acid for both ribosomal protein synthesis and cell growth.

...

[0126] As described above and below, the invention provides for nucleic acid polynucleotide sequences and polypeptide amino acid sequences, e.g., leucyl O-tRNAs and leucyl O-RSs, and, e.g., compositions, systems and methods comprising said sequences. Examples of said sequences, e.g., leucyl O-tRNAs and leucyl O-RSs are disclosed herein (see, Table 3, e.g., SEQ ID NO. 1-7, 12-16). However, one of skill in the art will appreciate that the invention is not limited to those sequences disclosed herein, e.g., as in the Examples. One of skill will appreciate that the invention also provides many related and unrelated sequences with the functions described herein, e.g., encoding a leucyl O-tRNA or a leucyl O-RS.

The invention provides polypeptides (leucyl O-RSs) and polynucleotides, 10 [0127]e.g., leucyl O-tRNA, polynucleotides that encode leucyl O-RSs or portions thereof, oligonucleotides used to isolate aminoacyl-tRNA synthetase clones, etc. Polynucleotides of the invention include those that encode proteins or polypeptides of interest of the invention with one or more selector codon. In addition, polynucleotides of the invention include, e.g., a polynucleotide comprising a nucleotide sequence as set forth in any one of SEQ ID NO.: 15 1-2, 4-7 and 12; a polynucleotide that is complementary to or that encodes a polynucleotide sequence thereof. A polynucleotide of the invention also includes a polynucleotide that encodes a polypeptide of the invention. Similarly, a nucleic acid that hybridizes to a polynucleotide indicated above under highly stringent conditions over substantially the entire length of the nucleic acid is a polynucleotide of the invention. In one embodiment, a 20 composition includes a polypeptide of the invention and an excipient (e.g., buffer, water, pharmaceutically acceptable excipient, etc.). The invention also provides an antibody or antisera specifically immunoreactive with a polypeptide of the invention.

[0128] A polynucleotide of the invention also includes a polynucleotide that is, e.g., at least 75%, at least 80%, at least 90%, at least 95%, at least 98% or more identical to that of a naturally occurring leucyl tRNA and comprises an anticodon loop comprising a CU(X) n XXXAA sequence, an stem region lacking noncanonical base pairs and a conserved discriminator base at position 73. A polynucleotide also includes a polynucleotide that is, e.g., at least 75%, at least 80%, at least 90%, at least 95%, at least 98% or more identical to that of a naturally occurring leucyl tRNA and comprises an anticodon loop comprising a CUUCCUAA sequence, a first pair selected from T28:A42, G28:C42 and/or C28:G42, and

25

30

a second pair selected from G:49:C65 or C49:G65, wherein the numbering corresponds to that indicated in Figure 4, Panel A.

[0129] In certain embodiments, a vector (e.g., a plasmid, a cosmid, a phage, a virus, etc.) comprises a polynucleotide of the invention. In one embodiment, the vector is an expression vector. In another embodiment, the expression vector includes a promoter operably linked to one or more of the polynucleotides of the invention. In another embodiment, a cell comprises a vector that includes a polynucleotide of the invention.

[0130] One of skill will also appreciate that many variants of the disclosed sequences are included in the invention. For example, conservative variations of the disclosed sequences that yield a functionally identical sequence are included in the invention. Variants of the nucleic acid polynucleotide sequences, wherein the variants hybridize to at least one disclosed sequence, are considered to be included in the invention. Unique subsequences of the sequences disclosed herein, as determined by, e.g., standard sequence comparison techniques, are also included in the invention.

Conservative variations

10

15

20

25

30

Sec. 35:

[0131] Owing to the degeneracy of the genetic code, "silent substitutions" (i.e., substitutions in a nucleic acid sequence which do not result in an alteration in an encoded polypeptide) are an implied feature of every nucleic acid sequence which encodes an amino acid. Similarly, "conservative amino acid substitutions," in one or a few amino acids in an amino acid sequence are substituted with different amino acids with highly similar properties, are also readily identified as being highly similar to a disclosed construct. Such conservative variations of each disclosed sequence are a feature of the present invention.

(typically less than 5%, more typically less than 4%, 2% or 1%) in an encoded sequence are "conservatively modified variations" where the alterations result in the deletion of an amino acid, addition of an amino acid, or substitution of an amino acid with a chemically similar amino acid. Thus, "conservative variations" of a listed polypeptide sequence of the present

invention include substitutions of a small percentage, typically less than 5%, more typically less than 2% or 1%, of the amino acids of the polypeptide sequence, with a conservatively selected amino acid of the same conservative substitution group. Finally, the addition of sequences which do not essentially alter the encoded activity of a nucleic acid molecule, such as the addition of a non-functional sequence, is a conservative variation of the basic nucleic acid.

[0133] Conservative substitution tables providing functionally similar amino acids are well known in the art. The following sets forth example groups which contain natural amino acids that include "conservative substitutions" for one another.

10

15

20

5

Nonpolar and/or Aliphatic Side Chains	Polar, Uncharged Side Chains	Aromatic Side Chains	Positively Charged Side Chains	Negatively Charged Side Chains
Glycine Alanine Valine Leucine Isoleucine	Serine Threonine Cysteine Methionine Asparagine	Phenylalanine Tyrosine Tryptophan	Lysine Arginine Histidine	Aspartate Glutamate
Proline	Glutamine			-

Nucleic Acid Hybridization

[0134] Comparative hybridization can be used to identify nucleic acids of the invention, such as SEQ ID NO.: 1-2, 4-7 and 12, including conservative variations of nucleic acids of the invention, and this comparative hybridization method is a preferred method of distinguishing nucleic acids of the invention. In addition, target nucleic acids which hybridize to the nucleic acids represented by SEQ ID NO: 1-2, 4-7 and 12 under high, ultra-high and ultra-ultra high stringency conditions are a feature of the invention. Examples of such nucleic acids include those with one or a few silent or conservative nucleic acid substitutions as compared to a given nucleic acid sequence.

[0135] A test nucleic acid is said to specifically hybridize to a probe nucleic acid when it hybridizes at least ½ as well to the probe as to the perfectly matched complementary target, i.e., with a signal to noise ratio at least ½ as high as hybridization of the probe to the target under conditions in which the perfectly matched probe binds to the perfectly matched

complementary target with a signal to noise ratio that is at least about 5x-10x as high as that observed for hybridization to any of the unmatched target nucleic acids.

[0136] Nucleic acids "hybridize" when they associate, typically in solution.
Nucleic acids hybridize due to a variety of well characterized physico-chemical forces, such
5 - as hydrogen bonding, solvent exclusion, base stacking and the like. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993) Laboratory Techniques in Biochemistry and Molecular Biology--Hybridization with Nucleic Acid Probes part I chapter 2, "Overview of principles of hybridization and the strategy of nucleic acid probe assays," (Elsevier, New York), as well as in Ausubel, supra. Hames and Higgins (1995)
10 Gene Probes 1 IRL Press at Oxford University Press, Oxford, England, (Hames and Higgins 1) and Hames and Higgins (1995) Gene Probes 2 IRL Press at Oxford University Press, Oxford, England (Hames and Higgins 2) provide details on the synthesis, labeling, detection and quantification of DNA and RNA, including oligonucleotides.

[0137] An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is 50% formalin with 1 mg of heparin at 42°C, with the hybridization being carried out overnight. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (see, Sambrook, supra for a description of SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove background probe signal. An example low stringency wash is 2x SSC at 40°C for 15 minutes. In general, a signal to noise ratio of 5x (or higher) than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

15

20

25

30

[0133] "Stringent hybridization wash conditions" in the context of nucleic acid hybridization experiments such as Southern and northern hybridizations are sequence dependent, and are different under different environmental parameters. An extensive guide to the hybridization of nucleic acids is found in Tijssen (1993), *supra*. and in Hames and Higgins, 1 and 2. Stringent hybridization and wash conditions can easily be determined empirically for any test nucleic acid. For example, in determining highly stringent hybridization and wash conditions are gradually increased (e.g., by increasing temperature, decreasing salt concentration, increasing

detergent concentration and/or increasing the concentration of organic solvents such as formalin in the hybridization or wash), until a selected set of criteria are met. For example, the hybridization and wash conditions are gradually increased until a probe binds to a perfectly matched complementary target with a signal to noise ratio that is at least 5x as high as that observed for hybridization of the probe to an unmatched target.

[0139] "Very stringent" conditions are selected to be equal to the thermal melting point (T_m) for a particular probe. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the test sequence hybridizes to a perfectly matched probe. For the purposes of the present invention, generally, "highly stringent" hybridization and wash conditions are selected to be about 5° C lower than the T_m for the specific sequence at a defined ionic strength and pH.

10

15

30

"Ultra high-stringency" hybridization and wash conditions are those in which the stringency of hybridization and wash conditions are increased until the signal to noise ratio for binding of the probe to the perfectly matched complementary target nucleic acid is at least 10x as high as that observed for hybridization to any of the unmatched target nucleic acids. A target nucleic acid which hybridizes to a probe under such conditions, with a signal to noise ratio of at least ½ that of the perfectly matched complementary target nucleic acid is said to bind to the probe under ultra-high stringency conditions.

increasing the hybridization and/or wash conditions of the relevant hybridization assay. For example, those in which the stringency of hybridization and wash conditions are increased until the signal to noise ratio for binding of the probe to the perfectly matched complementary target nucleic acid is at least 10x, 20X, 50X, 100X, or 500X or more as high as that observed for hybridization to any of the unmatched target nucleic acids. A target nucleic acid which hybridizes to a probe under such conditions, with a signal to noise ratio of at least ½ that of the perfectly matched complementary target nucleic acid is said to bind to the probe under ultra-ultra-high stringency conditions.

[0142] Nucleic acids which do not hybridize to each other under stringent conditions are still substantially identical if the polypeptides which they encode are substantially identical. This occurs, e.g., when a copy of a nucleic acid is created using the maximum codon degeneracy permitted by the genetic code.

Unique subsequences

5

10

15

25

30

[0143] In one aspect, the invention provides a nucleic acid that comprises a unique subsequence in a nucleic acid selected from the sequences of leucyl O-tRNAs and leucyl O-RSs disclosed herein. The unique subsequence is unique as compared to a nucleic acid corresponding to any known leucyl O-tRNA or leucyl O-RS nucleic acid sequence.

Alignment can be performed using, e.g., BLAST set to default parameters. Any unique subsequence is useful, e.g., as a probe to identify the nucleic acids of the invention.

[0144] Similarly, the invention includes a polypeptide which comprises a unique subsequence in a polypeptide selected from the sequences of leucyl O-RSs disclosed herein. Here, the unique subsequence is unique as compared to a polypeptide corresponding to any of known polypeptide sequence.

The invention also provides for target nucleic acids which hybridizes under stringent conditions to a unique coding oligonucleotide which encodes a unique subsequence in a polypeptide selected from the sequences of leucyl O-RSs wherein the unique subsequence is unique as compared to a polypeptide corresponding to any of the control polypeptides (e.g., parental sequences from which synthetases of the invention were derived, e.g., by mutation). Unique sequences are determined as noted above.

Sequence comparison, identity, and homology

[0146] The terms "identical" or percent "identity," in the context of two or more nucleic acid or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same, when compared and aligned for maximum correspondence, as measured using one of the sequence comparison algorithms described below (or other algorithms available to persons of skill) or by visual inspection.

[0147] The phrase "substantially identical," in the context of two nucleic acids or polypeptides (e.g., DNAs encoding an leucyl O-tRNA or leucyl O-RS, or the amino acid sequence of an O-RS) refers to two or more sequences or subsequences that have at least about 60%, about 80%, about 90-95%, about 98%, about 99% or more nucleotide or amino acid residue identity, when compared and aligned for maximum correspondence, as measured using a sequence comparison algorithm or by visual inspection. Such "substantially identical" sequences are typically considered to be "homologous," without reference to actual ancestry. Preferably, the "substantial identity" exists over a region of the

sequences that is at least about 50 residues in length, more preferably over a region of at least about 100 residues, and most preferably, the sequences are substantially identical over at least about 150 residues, or over the full length of the two sequences to be compared.

[0148] For sequence comparison and homology determination, typically one

sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

[0149] Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), by the search for similarity method of Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (see generally, Ausubel et al., infra).

15

20

25

30

sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al.*, J. Mol. Biol. 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov/). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to

telija.

10

calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915).

[0151] In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul, Proc. Nat'l. Acad. Sci. USA 90:5873-5787 (1993)). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

20 Mutagenesis and Other Molecular Biology Techniques [0152] Polynucleotide and polypeptides of the invention and used in the invention can be manipulated using molecular biological techniques. General texts which describe molecular biological techniques include Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA (Berger); Sambrook et al., Molecular Cloning - A Laboratory Manual (2nd Ed.), Vol. 1-3, 25 Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook") and Current Protocols in Molecular Biology, F.M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 1999) ("Ausubel")). These texts describe mutagenesis, the use of . 30 vectors, promoters and many other relevant topics related to, e.g., the generation of genes that include selector codons for production of proteins that include selected amino acids

(e.g., unnatural amino acids), leucyl orthogonal tRNAs, leucyl orthogonal synthetases, and pairs thereof.

[0153]Various types of mutagenesis are used in the invention, e.g., to mutate tRNA molecules, to produce libraries of leucyl tRNAs, to produce libraries of leucyl synthetases, 5 _ to insert selector codons that encode a selected amino acid in a protein or polypeptide of interest. They include but are not limited to site-directed, random point mutagenesis, homologous recombination, DNA shuffling or other recursive mutagenesis methods, chimeric construction, mutagenesis using uracil containing templates, oligonucleotidedirected mutagenesis, phosphorothioate-modified DNA mutagenesis, mutagenesis using 10 gapped duplex DNA or the like, or any combination thereof. Additional suitable methods include point mismatch repair, mutagenesis using repair-deficient host strains, restrictionselection and restriction-purification, deletion mutagenesis, mutagenesis by total gene synthesis, double-strand break repair, and the like. Mutagenesis, e.g., involving chimeric constructs, is also included in the present invention. In one embodiment, mutagenesis can 15 be guided by known information of the naturally occurring molecule or altered or mutated

naturally occurring molecule, e.g., sequence, sequence comparisons, physical properties,

crystal structure or the like.

20

25

30

[0154] Host cells are genetically engineered (e.g., transformed, transduced or transfected) with the polynucleotides of the invention or constructs which include a polynucleotide of the invention, e.g., a vector of the invention, which can be, for example, a cloning vector or an expression vector. For example, the coding regions for the orthogonal tRNA, the orthogonal tRNA synthetase, and the protein to be derivatized are operably linked to gene expression control elements that are functional in the desired host cell. Typical vectors contain transcription and translation terminators, transcription and translation initiation sequences, and promoters useful for regulation of the expression of the particular target nucleic acid. The vectors optionally comprise generic expression cassettes containing at least one independent terminator sequence, sequences permitting replication of the cassette in eukaryotes, or prokaryotes, or both (e.g., shuttle vectors) and selection markers for both prokaryotic and eukaryotic systems. Vectors are suitable for replication and/or integration in prokaryotes, eukaryotes, or preferably both. See, Giliman & Smith, Gene 8:81 (1979); Roberts, et al., Nature, 328:731 (1987); Schneider, B., et al., Protein Expr. Purif. 6435:10 (1995); Ausubel, Sambrook, Berger (all supra). The vector can be, for

....

10

15

example, in the form of a plasmid, a bacterium, a virus, a naked polynucleotide, or a conjugated polynucleotide. The vectors are introduced into cells and/or microorganisms by standard methods including electroporation (From et al., <u>Proc. Natl. Acad. Sci. USA</u> 82, 5824 (1985), infection by viral vectors, high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface (Klein et al., <u>Nature</u> 327, 70-73 (1987)), and/or the like.

[0155] A catalogue of Bacteria and Bacteriophages useful for cloning is provided, e.g., by the ATCC, e.g., The ATCC Catalogue of Bacteria and Bacteriophage (1992)

Gherna et al. (eds) published by the ATCC. Additional basic procedures for sequencing, cloning and other aspects of molecular biology and underlying theoretical considerations are also found in Watson et al. (1992) Recombinant DNA Second Edition Scientific American Books, NY. In addition, essentially any nucleic acid (and virtually any labeled nucleic acid, whether standard or non-standard) can be custom or standard ordered from any of a variety of commercial sources, such as the Midland Certified Reagent Company (Midland, TX mcrc.com), The Great American Gene Company (Ramona, CA available on the World Wide Web at genco.com), ExpressGen Inc. (Chicago, IL available on the World Wide Web at expressgen.com), Operon Technologies Inc. (Alameda, CA) and many others.

[0156] The engineered host cells can be cultured in conventional nutrient media modified as appropriate for such activities as, for example, screening steps, activating promoters or selecting transformants. These cells can optionally be cultured into transgenic organisms. Other useful references, e.g. for cell isolation and culture (e.g., for subsequent nucleic acid isolation) include Freshney (1994) Culture of Animal Cells, a Manual of Basic Technique, third edition, Wiley- Liss, New York and the references cited therein; Payne et al. (1992) Plant Cell and Tissue Culture in Liquid Systems John Wiley & Sons, Inc. New York, NY; Gamborg and Phillips (eds) (1995) Plant Cell, Tissue and Organ Culture; Fundamental Methods Springer Lab Manual, Springer-Verlag (Berlin Heidelberg New York) and Atlas and Parks (eds) The Handbook of Microbiological Media (1993) CRC Press, Boca Raton, FL.

PROTEINS AND POLYPEPTIDES OF INTEREST

30 [0157] Methods of producing a protein in a cell with a selected amino acid at a specified position are also a feature of the invention. For example, a method includes

growing, in an appropriate medium, the cell, where the cell comprises a nucleic acid that comprises at least one selector codon and encodes a protein; and, providing the selected amino acid; where the cell further comprises: an orthogonal leucyl-tRNA (leucyl-O-tRNA) that functions in the cell and recognizes the selector codon; and, an orthogonal aminoacyl-tRNA synthetase (O-RS) that preferentially aminoacylates the leucyl-O-tRNA with the selected amino acid. Typically, the leucyl-O-tRNA comprises at least about a 25%, 50%, 75%, 80%, 85%, 90%, 95% or 98% suppression activity in the presence of a cognate synthetase in response to a selector codon as compared to a control lacking the selector codon (and, typically, the cognate synthetase). A protein produced by this method is also a feature of the invention.

- [0158] The invention also teaches variant orthogonal leucyl-tRNAs and variant orthogonal aminoacyl-tRNA synthetase species that display suppression activity, where the suppression activity is measured relative to the suppression activity of a leucyl-O-tRNA nucleotide sequence or an O-RS amino acid sequence provided by the present invention.

 For example, the invention teaches variant leucyl-O-tRNA species that display suppression activity that is at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% as effective as a leucyl-O-tRNA sequence provided by the examples herein (e.g., SEQ ID NOs: 1-7 and 12). Similarly, the invention teaches variant O-RS species that display suppression activity that is at least 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 98% or 99% as effective as an O-RS sequence provided by the examples herein (e.g., SEQ ID NO: 15 and 16).
 - [0159] In another aspect, the invention teaches variant orthogonal leucyl-tRNAs and variant orthogonal aminoacyl-tRNA synthetase species that display suppression activity that is equal to or greater than the suppression activity of a leucyl-O-tRNA nucleotide sequence or an O-RS amino acid sequence provided by the present specification. For example, the invention teaches variant leucyl-O-tRNA species that display suppression activity that is at least 100% as effective as a leucyl-O-tRNA sequence provided by the examples herein (e.g., SEQ ID NO: 1-7 and 12).

25

30

[0160] The compositions of the invention and compositions made by the methods of the invention optionally are in a cell. The leucyl O-tRNA/O-RS pairs or individual components of the invention can then be used in a host system's translation machinery, which results in a selected amino acid, e.g., unnatural amino acid, being incorporated into a

A service

protein. The International Application Number PCT/US2004/011786, filed April 16, 2004, and WO 2002/085923, entitled "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS" describe this process, and is incorporated herein by reference. For example, when an leucyl O-tRNA/O-RS pair is introduced into a host, e.g., Escherichia coli, the pair leads to the in vivo incorporation of selected amino acid, such as an unnatural amino acid, e.g., a synthetic amino acid, such as derivative of a leucine amino acid, which can be exogenously added to the growth medium, into a protein, in response to a selector codon. Optionally, the compositions of the present invention can be in an in vitro translation system, or in an in vivo system(s).

- 10 [0161] Essentially any protein (or portion thereof) that includes a selected amino acid, e.g., an unnatural amino acid, (and any corresponding coding nucleic acid, e.g., which includes one or more selector codons) can be produced using the compositions and methods herein. No attempt is made to identify the hundreds of thousands of known proteins, any of which can be modified to include one or more unnatural amino acid, e.g., by tailoring any available mutation methods to include one or more appropriate selector codon in a relevant translation system. Common sequence repositories for known proteins include GenBank EMBL, DDBJ and the NCBI. Other repositories can easily be identified by searching the internet.
- [0162] Typically, the proteins are, e.g., at least 60%, at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, or at least 99% or more identical to any available protein (e.g., a therapeutic protein, a diagnostic protein, an industrial enzyme, or portion thereof, and the like), and they comprise one or more selected amino acid. Examples of therapeutic, diagnostic, and other proteins that can be modified to comprise one or more selected amino acid, e.g., an unnatural amino acid, can be found, but not limited to, those in International Application Number PCT/US2004/011786, filed April 16, 2004, and WO 2002/085923, entitled "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS."
- [0163] In certain embodiments, the protein or polypeptide of interest (or portion thereof) in the methods and/or compositions of the invention is encoded by a nucleic acid.
 Typically, the nucleic acid comprises at least one selector codon, at least two selector codons, at least three selector codons, at least five selector

codons, at least six selector codons, at least seven selector codons, at least eight selector codons, at least nine selector codons, ten or more selector codons.

[0164] Genes coding for proteins or polypeptides of interest can be mutagenized using methods well-known to one of skill in the art and described herein under

- "Mutagenesis and Other Molecular Biology Techniques" to include, e.g., one or more selector codon for the incorporation of a selected amino acid, e.g., an unnatural amino acid. For example, a nucleic acid for a protein of interest is mutagenized to include one or more selector codon, providing for the insertion of the one or more selected amino acids, e.g., unnatural amino acids. The invention includes any such variant, e.g., mutant, versions of any protein, e.g., including at least one selected amino acid. Similarly, the invention also includes corresponding nucleic acids, i.e., any nucleic acid with one or more selector codon that encodes one or more selected amino acid.
 - [0165] To make a protein that includes a selected amino acid, one can use host cells and organisms that are adapted for the in vivo incorporation of the selected amino acid via orthogonal leucyl tRNA/RS pairs. Host cells are genetically engineered (e.g., transformed, transduced or transfected) with one or more vectors that express the orthogonal leucyl tRNA, the orthogonal leucyl tRNA synthetase, and a vector that encodes the protein to be derivatized. Each of these components can be on the same vector, or each can be on a separate vector, two components can be on one vector and the third component on a second vector. The vector can be, for example, in the form of a plasmid, a bacterium, a virus, a naked polynucleotide, or a conjugated polynucleotide.

Defining Polypeptides by Immunoreactivity

15

20

25

30

[0166] Because the polypeptides of the invention provide a variety of new polypeptide sequences (e.g., comprising selected amino acids (e.g., unnatural amino acids) in the case of proteins synthesized in the translation systems herein, or, e.g., in the case of the novel synthetases, novel sequences of standard amino acids), the polypeptides also provide new structural features which can be recognized, e.g., in immunological assays. The generation of antisera, which specifically bind the polypeptides of the invention, as well as the polypeptides which are bound by such antisera, are a feature of the invention. The term "antibody," as used herein, includes, but is not limited to a polypeptide substantially encoded by an immunoglobulin gene or immunoglobulin genes, or fragments thereof which specifically bind and recognize an analyte (antigen). Examples include polyclonal,

monoclonal, chimeric, and single chain antibodies, and the like. Fragments of immunoglobulins, including Fab fragments and fragments produced by an expression library, including phage display, are also included in the term "antibody" as used herein. See, e.g., Paul, Fundamental Immunology, 4th Ed., 1999, Raven Press, New York, for antibody structure and terminology.

[0167] In order to produce antisera for use in an immunoassay, one or more of the immunogenic polypeptides is produced and purified as described herein. For example, recombinant protein can be produced in a recombinant cell. An inbred strain of mice (used in this assay because results are more reproducible due to the virtual genetic identity of the mice) is immunized with the immunogenic protein(s) in combination with a standard adjuvant, such as Freund's adjuvant, and a standard mouse immunization protocol (see, e.g., Harlow and Lane (1988) Antibodies, A Laboratory Manual, Cold Spring Harbor Publications, New York, for a standard description of antibody generation, immunoassay formats and conditions that can be used to determine specific immunoreactivity.

15 [0168] Additional details on proteins, antibodies, antisera, etc. can be found in, e.g., International Application Number PCT/US2004/011786, filed April 16, 2004, WO 2002/085923, entitled "IN VIVO INCORPORATION OF UNNATURAL AMINO ACIDS;" International Application Numbers PCT/US2003/32870, filed October 15, 2003; and PCT/US2003/41346, filed December 22, 2003.

20 <u>KITS</u>

25

10

[0169] Kits are also a feature of the invention. For example, a kit for producing a protein that comprises at least one selected amino acid, e.g., an unnatural amino acid, in a cell is provided, where the kit includes a container containing a polynucleotide sequence encoding an leucyl O-tRNA, and/or an leucyl O-tRNA, and/or a polynucleotide sequence encoding an leucyl O-RS, and/or an leucyl O-RS. In one embodiment, the kit further includes at least selected amino acid. In another embodiment, the kit further comprises instructional materials for producing the protein.

EXAMPLES

[0170] The following examples are offered to illustrate, but not to limit the claimed invention. One of skill will recognize a variety of non-critical parameters that may be altered without departing from the scope of the claimed invention.

EXAMPLE 1: ADAPTATION OF AN ORTHOGONAL ARCHAEAL LEUCYL-TRNA AND SYNTHETASE PAIR FOR FOUR-BASE, AMBER, AND OPAL SUPPRESSION

[0171]Recently, it has been shown that an amber suppressor tRNA-aminoacyl tRNA synthetase pair derived from the tyrosyl-tRNA synthetase of Methanococcus jannaschii can be used to genetically encode unnatural amino acids in response to the amber nonsense codon, TAG. This pair is unable to decode either the opal nonsense codon, TGA, or the four-base codon, AGGA. To overcome this, a leucyl-tRNA synthetase from Methanobacterium thermoautotrophicum and leucyl tRNA derived from Halobacterium sp. NRC-1 was adapted as an orthogonal tRNA-synthetase pair in E. coli to decode amber 10 (TAG), opal (TGA) and four-base (AGGA) codons. To improve the efficiency and selectivity of the suppressor tRNA, extensive mutagenesis was performed on the anticodon loop and acceptor stem. The two most significant criteria required for an efficient amber orthogonal suppressor tRNA are a CU(X)XXXAA anticodon loop and the lack of noncanonical or mismatched base pairs in the stem regions. These changes afford only weak suppression of TGA and AGGA. However, this information, together with an analysis of 15 sequence similarity of multiple native archaeal tRNA sequences, led to efficient, orthogonal suppressors of opal codons and the four-base codon, AGGA. Ultimately, these additional orthogonal pairs can be used to genetically incorporate multiple unnatural amino acids into

20 [0172] A great deal of effort has focused on the cotranslational incorporation of unnatural amino acids into proteins. Early work demonstrated that the translational machinery of E. coli would accommodate amino acids similar in structure to the common twenty (Hortin and Boime (1983) Methods Enzymol. 96, 777-784). This work was further extended by relaxing the specificity of endogenous E. coli synthetases so that they activate unnatural amino acids as well as their cognate natural amino acid. Moreover, it was shown that mutations in editing domains could also be used to extend the substrate scope of the endogenous synthetase (Doring et al., (2001) Science 292, 501-504). However, these strategies are limited to recoding the genetic code rather than expanding the genetic code and lead to varying degrees of substitution of one of the common twenty amino acids with an unnatural amino acid.

proteins.

[0173] Later it was shown that unnatural amino acids could be site-specifically incorporated into proteins *in vitro* by the addition of chemically aminoacylated orthogonal amber

suppressor tRNAs to an in vitro transcription/translation reaction (Noren et al., (1989) Science 244, 182-188; Bain et al., (1989) J. Am. Chem. Soc. 111, 8013-8014; Dougherty (2000) Curr. Opin. Chem. Biol. 4, 645-652; Cornish et al., (1995) Angew. Chem., Int. Ed. 34, 621-633). It is clear from these studies that the ribosome and translation factors are compatible with a large number of unnatural amino acids, even those with unusual structures. Unfortunately, the chemical aminoacylation of tRNAs is difficult, and this method can only produce microgram-scale quantities of protein due to the stoichiometric nature of the process. A catalytic in vivo method could overcome these limitations, and would also permit the study of proteins containing unnatural amino acids in living cells. [0174] In order to add additional synthetic amino acids to the genetic code in vivo it is 10 necessary to generate a 21st "orthogonal pair" of synthetase and tRNA that can function efficiently in the translational machinery. The synthetase should not cross-react with any of the endogenous tRNAs (40 in E. coli), and the orthogonal tRNA should not be aminoacylated by any of the endogenous synthetases (21 in E. coli). The tRNA should 15 decode only a specific new codon that is not decoded by any endogenous tRNA, and the synthetase should charge its tRNA with only a specific unnatural amino acid. We have successfully generated an orthogonal tRNA-synthetase pair from tyrosyl-tRNA synthetase of Methanococcus jannaschii which satisfies these requirements. This system has been used to incorporate a series of unnatural amino acids including keto amino acids (Wang et 20 al., (2003) Proc. Natl. Acad. Sci. U. S. A. 100, 56-61), photocrosslinking amino acids (Chin et al., (2002) Proc. Natl. Acad. Sci. U. S. A. 99, 11020-11024; Chin et al., (2002) J. Am. Chem. Soc. 124, 9026-9027), and heavy atom containing amino acids selectively into proteins in response to the TAG codon. [0175] Several other orthogonal pairs have been reported. Glutaminyl (Liu and Schultz 25 (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 4780-4785), aspartyl (Pastrnak et al., (2000) Helv.

(1999) Proc. Natl. Acad. Sci. U. S. A. 96, 4780-4785), aspartyl (Pastrnak et al., (2000) Helv. Chim. Acta 83, 2277-2286), and tyrosyl (Ohno et al., (1998) J. Biochem. (Tokyo, Jpn.) 124, 1065-1068; Kowal et al., (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 2268-2273) systems derived from S. cerevisiae tRNAs and synthetases have been described for the potential incorporation of unnatural amino acids in E. coli. Systems derived from the E. coli glutaminyl (Kowal et al., (2001) Proc. Natl. Acad. Sci. U. S. A. 98, 2268-2273) and tyrosyl (Edwards and Schimmel (1990) Mol. Cell. Biol. 10, 1633-1641) synthetase have been described for use in S. cerevisiae. The E. coli tyrosyl system can also function in mammalian cells and has been used for the incorporation of 3-iodo-L-tyrosine in vivo

30

(Sakamoto et al., (2002) Nucleic Acids Res. 30, 4692-4699). All of these systems have made exclusive use of the amber stop codon. To expand the genetic code beyond twentyone amino acids, other orthogonal pairs and unique codons need to be identified. [0176] A desired property of any orthogonal pair are a codon that is unique within the genetic code and that will not cross-react with noncognate tRNAs. In addition to the amber 5 stop codon (TAG), the opal nonsense codon (TGA) is one such candidate. A genetic code in which TAG and TGA encoded unnatural amino acids could encode 22 amino acids while preserving the ochre nonsense codon, UAA, which is the most abundant termination signal. The suppression of opal codons is robust in vivo but has not been frequently used for the 10 incorporation of unnatural amino acids in vitro due to high background readthrough of TGA codons (Cload et al., (1996) Chem. Biol. 3, 1033-1038). Another possible codon involves unnatural base pairs. Unnatural amino acids have been incorporated in response to novel codons containing the unnatural base (iso-dC)AG (Piccirilli et al., (1990) Nature 343, 33-37) or pyridin-2-one (Hirao et al., (2002) Nat. Biotechnol. 20, 177-182) using an in vitro 15 translation system. Adaptation of unnatural base pairs for the incorporation of unnatural amino acids into proteins in vivo, need the faithful replication and transcription of unnatural base pairs in DNA and RNA (Wu et al., (2002) J. Am. Chem. Soc. 124, 14626-14630). Another codon that can used to encode additional amino acids are four- and five-base codons. Using a library of tRNAs with randomized anticodon loops coupled with a - 20 selection scheme, several highly efficient and non cross-reactive four- and five-base codons. were identified, including AGGA, UAGA, CCCU, and UAGA (Magliery et al., (2001) J. Mol. Biol. 307, 755-769; Anderson et al., (2002) Chem. Biol. 9, 237-244). [0177] Regardless of the codon chosen, it is useful to generate additional orthogonal tRNAsynthetase pairs that can translate these codons with high fidelity and good efficiency. 25 Because the tRNA anticodon loop is a major identity element for recognition by most synthetases, one must identify a synthetase that does not recognize these identity elements in order to generate suppressor tRNAs for these unusual codons. The leucyl-, seryl-, and alanyl-tRNA synthetases of E. coli are well known to tolerate extensive substitutions in the anticodon loop (Shimizu et al., (1992) J. Mol. Evol. 35, 436-443; Kleina (1990) J. Mol. 30 Biol. 213, 705-717; Sampson and Saks (1993) Nucleic Acids Res. 21, 4467-4475). Some homologous archaeal or eukaryotic synthetases may have similar properties. Herein are derivatives of a leucyl-tRNA synthetase from Methanobacterium thermoautotrophicum and

leucyl tRNAs derived from Halobacterium sp. NRC-1 that act as orthogonal tRNA-

synthetase pairs for the amber codon in *E. coli*. Moreover, information gained in these studies, together with multiple sequence alignments of native archaeal tRNA sequences, allowed us to design efficient orthogonal suppressor tRNAs of opal codons and a four-base codon, AGGA.

Material and Methods

5

10

15

1. 4. 3.

[0178] Strains, plasmids, and materials. All in vivo manipulations were carried out in E. coli strain TOP10 (Invitrogen) in LB media at 37°C. Halobacterium sp. NRC-1 was purchased from the American Type Culture Collection (ATCC). PCR was carried out according to standard protocols with a mixture of Taq (Promega) and Pfu (Stratagene) polymerases. Oligonucleotides were synthesized by Genosys, Operon, or the UCSF Biomolecular Resource Center. For oligonucleotides containing degenerate bases, the phosphoramidites were premixed to avoid bias. Standard protocols were employed for subcloning with restriction enzymes (NEB) and T4 DNA ligase (NEB). Plasmids were introduced into E. coli by electroporation. Sequence analysis was performed using the Genetics Computer Group, Inc. (GCG) software. The sequences of all plasmids were confirmed by restriction mapping and sequencing.

from ATCC or was prepared from a cell pellet purchased from ATCC. Genomic DNA was extracted using the DNeasy kit (Qiagen). Synthetase genes were amplified from genomic DNA by PCR then subcloned into the NcoI and either EcoRI, KpnI, or PvuII sites of plasmid pKQ. More details on the cloning of these genes can be found Table 2 and is also available on the Internet at http://pubs.acs.org. Plasmid pKQ contains the ribosome binding site, multiple cloning site, and rnnB terminator from plasmid pBAD-Myc/HisA (Invitrogen) under control of a constitutive glutamine promoter. The plasmid also contains a CoIE1 origin of replication, and a kanamycin resistance gene for plasmid maintenance.

[0180] Constructions of reporter plasmids. Beta-lactamase reporter plasmids were constructed from plasmid pACKO-Bla. This plasmid was constructed with a p15a origin, a chloramphenicol resistance gene, and unique sites for insertion of a gene for β-lactamase and a tRNA under control of the strong, constitutive *lpp* promoter. Site A184 of the β-lactamase gene was changed to TAG, AGGA, or TGA by an overlap PCR strategy, and the genes were subcloned into the *Aat*II and *Xma*I sites of pACKO-Bla to give plasmids pACKO-A184TAG, pACKO-A184AGGA, and pACKO-A184TGA.

[0181] Constructions of tRNA plasmids. Genes for individual tRNAs and for tRNA libraries were constructed by extension reactions and subcloned into the EcoRI and PstI sites of pACKO-Bla derivatives. All libraries represented at least 10-fold more members than the theoretical size of the library to ensure high coverage.

- 5 [0182] Measurement of suppression efficiency. A series of LB agar plates were prepared with 25 μg/mL of kanamycin, 25 μg/mL of chloramphenicol, and concentrations of ampicillin between 5 and 1000 μg/mL. Synthetase and tRNA plasmids were cotransformed and plated at densities below 100 cells per plate. Suppression efficiency was reported as the highest concentration at which cells survived to form colonies among a series of plates for which the next highest and lowest concentrations would be within 20% of the reported value.
- [0183] Selection of libraries and characterization of selectants. All tRNA libraries were subjected to ampicillin selection and the surviving colonies were isolated and sequenced by the method described previously (Magliery et al., (2001) J. Mol. Biol. 307, 755-769). Briefly, libraries were spread on LB plates containing 25 μg/mL of kanamycin and chloramphenicol for plasmid maintenance, and varying concentrations of ampicillin for selection. After 24 hours of growth, the plates were scraped, and the cells were diluted slightly then spread again on ampicillin plates. After colonies appeared, plates were again scraped and plated at dilute cell densities on a range of plates with different ampicillin concentrations. Selectants were isolated, sequenced, and then confirmed by retransformation into cells containing synthetase-expressing plasmids.
- [0184] Beta-galactosidase reporter assays. The full-length lacZ gene of plasmid pBAD-Myc/His/lacZ (Invitrogen) was amplified by PCR and subcloned into plasmid pLASC to obtain plasmid pLASC-lacZ. This pSC101-derived plasmid expresses lacZ gene under the control of an lpp promoter and has an ampicillin resistance gene for plasmid maintenance. Derivatives of this plasmid were constructed wherein Leu-25 of the peptide VVLQRRDWEN of lacZ was replaced by TAG, TGA, or AGGA codons, or sense codons for tyrosine, serine, or leucine. The appropriate pLASC-lacZ-, pACKO-Bla-, and pKQ-derived plasmids were cotransformed and grown to an OD₆₀₀ of 0.5. Beta-galactosidase assays were performed in quadruplicate using the BetaFluorTM β-Galactosidase Assay Kit (Novagen). Percent suppression was calculated as the percentage of activity for a sample

relative to the value observed from the pLASC-lacZ construct with the corresponding sense codon at position 25. Cells containing pLASC-lacZ plasmids with sense codons at position 25 were also assayed by 2-nitrophenyl-β-D-galactopyranoside assays (Miller (1972) Experiments in molecular genetics, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.), and activity was calculated in Miller units.

[0185] Purification of synthetase proteins. Synthetase genes were cloned in frame with the C-terminal myc/his tag of pBAD-Myc/HisA (Invitrogen). Protein purification was performed with the Qiaexpressionist kit (Qiagen) by the manufacturer's protocol under native conditions. Protein concentrations were measured by the BCA Protein Assay Kit (Pierce) and analyzed by SDS-PAGE.

10

[0186] In vitro aminoacylation assays. Aminoacylation assays were performed by methods described previously (Hoben and Soll (1985) Methods Enzymol. 113, 55-59) in 20 μL reactions containing 50 mM Tris-Cl, pH 7.5, 30 mM KCl, 20 mM MgCl₂, 3 mM glutathione, 0.1 mg/mL BSA, 10 mM ATP, 1 μM (79 Ci/mmol) [³H] leucine (Amersham),

750 nM synthetase, and 0, 2, 10, or 40 μM crude total tRNA. Crude total *E. coli* tRNA was purchased from Roche, and halobacterial tRNA was extracted from cultures of *Halobacterium sp. NRC-1* with the RNA/DNA Extraction Kit (Qiagen).

[0187] Detailed information for the cloning of archaeal leucyl-tRNA synthetases can be found below and is also available on the Internet at http://pubs.acs.org.

20 TABLE 2: CLONING OF ARCHAEAL LEUCYL-tRNA SYNTHETASES
INTO PLASMID pKO

	Acception	Famous	T 5		
Organism	Accession number	ATCC Number	Forward Oligo	Reverse Oligo	
Halobacterium sp. NRC-1	NP_280869.1	700922	ca214	ca215	
Escherichia coli (strain HB101)	P07813	N/A	ca244	ca215	
Methanococcus jannaschii	Q58050	43067D	ca246	ca247	
Archaeoglobus fulgidus	O30250	49558D	ca261	ca247	
Aeropyrum pernix K1	Q9YD97	700793D	ca263	ca264	
Pyrococcus horikoshii	O58698	700860	ca265	ca266	
Methanobacterium				ca275	
thermoautotrophicum	027552	700791	ca274	1	

Organism	Restriction enzymes		
Halobacterium sp. NRC-1	Ncol/EcoRI		
Escherichia coli (strain HB101)	Ncol/EcoRI		
Methanococcus jannaschii	Ncol/Kpnl		
Archaeoglobus fulgidus	Ncol/Kpnl		

Aeropyrum pernix K1	BspHI/EcoRI (subcloned into Ncol/EcoRI sites)
Pyrococcus horikoshii	Ncol/EcoRV (subcloned into Ncol/Pvull sites)
Methanobacterium thermoautotrophicum	BsmBl (subcloned into Ncol/EcoRl sites)

Oligo Sequences

5

10

15

20

	Ongo Ser	quences
	ca214	GGTTTCCATGGGAGAGCAAGCCACCTAC
-	ca215	GGTTTGGAATTCAGTCGTCGGCTTCGTCG
	ca244	CGAAACCATGGAAGAGCAATACCGCCCGGAAG
	ca245	CCAAAGAATTCCCGCCAACGACCAGATTGAGGAG
	ca246	CGAAACCATGGTTATGATTGACTTTAAAG
	ca247	CGAAAGGTACCTTGTATTCAAGATAAATAGCTGG
i	ca261	GCGAACCATGGGCGATTTCAGGATAATTGAG
	ca262	CAATTGGTACCTTAAGCAACATAAATCGCG
	ca263	GGATTATCATGAAGCGACTAAAGGCCGTGGAGGAG
	ca264	CACTTGAATTCTTAGCCTCCTCTCTCTCCGC
	ca265	CGAATCCATGGCTGAGCTTAACTTCAAGG
	ca266	GGATGGATATCACTCGATGAAGATGGCAG
ļ	ca274	GGAGACGTCTCTCATGGATATTGAAAGAAAATGGCG
	ca275	CGTTACGTCTCGAATTGGAAAAGAGCTGTCTGAGG

Results and Discussion

[0188] Identification of orthogonal tRNAs. Previous studies (Kwok and Wong (1980) Can. J. Biochem. 58, 213-218) have shown that halobacterial tRNAs are inefficiently charged by the E. coli leucyl-tRNA synthetase. The similarity between the halobacterial and other archaeal leucyl-tRNAs (see Figure 1, Panel C) led us to believe that tRNAs from other archaeans can also be orthogonal to the E. coli synthetases. The sequences were chosen to broadly represent the family of archaeal leucyl-tRNAs and included tRNA₃^{Leu} of Archaeoglobus fulgidus (AfL3), tRNA₄^{Leu} of Halobacterium sp. NRC-1 (HhL4), tRNA₂^{Leu} of Methanococcus jannaschii (MjL2), tRNA₅^{Leu} of Pyrococcus furiosus (PfL5), and tRNA₂^{Leu} of Fyrococcus horikoshi (PhL2) (see Figure 1, Panel C for sequences). In all cases, the anticodon was changed to CUA, and CCA was added to the 3' terminus if the sequence was not present in the source gene to obtain an amber suppressor tRNA.

[0189] To measure the activity of suppressor tRNAs, a selection system was developed based on the *in vivo* suppression of nonsense or frameshift mutations introduced into the gene for β-lactamase (*bla*). Reporter genes for *bla* variants with TAG, AGGA, and TGA at position A184 (a permissive site (Liu and Schultz (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 4780-4785)) were constructed in plasmid pACKO-Bla, a medium copy plasmid

4.

10

15

derived from pACYC184. Bacteria transformed with these reporter constructs are unable to grow on LB agar plates with ampicillin concentrations greater than 5 μg/mL, only slightly higher than the value (2 μg/mL ampicillin) observed for bacteria transformed with no plasmids. Plasmids derived from pACKO-Bla can also express tRNA genes under the control of a strong *lpp* promoter. When the robust amber suppressor gene *supD*, a tRNA efficiently charged by *E. coli* seryl tRNA synthetase, is expressed from pACKO-A184TAG (which encodes the A184TAG variant of *bla*), host bacteria survive at an ampicillin concentration of 1000 μg/mL. In contrast, in the case of an orthogonal tRNA, which cannot be efficiently charged by endogenous *E. coli* synthetases, ampicillin resistance should be less than 5 ug/mL. Conversely, if the tRNA is not orthogonal, or if a heterologous synthetase capable of charging the tRNA is co-expressed in the system, a higher level of ampicillin resistance should be observed.

[0190] The genes for the five potential orthogonal amber suppressor tRNAs were integrated into pACKO-A184TAG. *E. coli* hosts expressing the HhL4-derived suppressor, designated HL(TAG)1, could survive to only 5 μ g/mL ampicillin, the MjL2- and PhL2-derived suppressors to 7 μ g/mL, and the PfL5- and AfL3-derived suppressors to 20 μ g/mL ampicillin. Therefore, all five suppressor tRNAs are either weak suppressor tRNAs or are inefficiently charged by *E. coli* aminoacyl-tRNA synthetases.

[0191] Cloning of archaeal leucyl-tRNA synthetases. Due to the high homology of the archaeal leucyl-tRNAs, we anticipated that the archeal leucyl-tRNA synthetases might have similar tRNA recognition properties. Therefore both species specific and cross species combinations of archeal leucyl-tRNAs and synthetases were examined in order to find an optimal pair for use in E. coli. The leucyl-tRNA synthetases from Archaeoglobus fulgidus (AfLRS), Aeuropyrum pernix (ApLRS), Halobacterium sp. NRC-1 (HhLRS),

Methanococcus jannaschii (MjLRS), Methanobacterium thermoautotrophicum (MtLRS), and Pyrococcus horikoshi (PhLRS) were chosen as initial candidates due to the availability of the genome sequences and commercial availability of the organisms. The genes for these synthetases were cloned under the control of a constitutive glutamine promoter on the high copy plasmid, pKQ, which was constructed from pBR322 and contains a kanamycin resistance gene. The leuS gene from E. coli (EcLRS) was also cloned as a negative control.

Synthetase expression plasmids and reporter constructs were cotransformed and assayed for activity by ampicillin selection (Figure 2). In general, the reporter plasmid containing the

41,000

10

15

20

25

30

HhL4 suppressor tRNA, HL(TAG)1, gave the largest enhancement in suppression efficiency upon cotransformation with synthetase-expressing plasmids, but the PhL2- and AfL3-derived tRNAs also show a suppression enhancement. The MjL2- and PfL5-derived suppressor tRNAs survive to the same concentrations of ampicillin regardless of whether or not the archaeal synthetase is present, and were not pursued further. From all 35 combinations of synthetase and reporter plasmids, the highest levels of ampicillin resistance result when the synthetases, MtLRS or MjLRS, are expressed with the HhL4-derived suppressor tRNA. The AfLRS construct gives slightly lower levels of resistance, and all other synthetases give no increase in suppression efficiency over background levels. With MjLRS or MtLRS, cells expressing HL(TAG)1 survive to 35 μg/mL ampicillin, but only 5 μg/mL with the *E. coli* synthetase or plasmid lacking the synthetase. Cells expressing AfLRS can survive to 25 μg/mL ampicillin when coexpressed with HL(TAG)1. From these *in vivo* suppression screens, three synthetases, MtLRS, MjLRS, and AfLRS, were identified as candidates for an orthogonal pair with the HhL4-derived amber suppressor tRNA, HL(TAG)1.

[0192] In vitro charging assays. An in vivo suppression screen can distinguish active and inactive aminoacyl-tRNA synthetases, but it cannot distinguish an orthogonal synthetase from one that cross-reacts with E. coli tRNA. To determine the permissiveness of AfLRS, MiLRS, and MtLRS for E. coli tRNA, the synthetases were overexpressed, purified, and then subjected to in vitro aminoacylation assays to measure their ability to charge E. coli tRNA. AfLRS, MiLRS, and MtLRS were purified from an arabinose promoter over-expression system by Ni-NTA affinity chromatography in yields of 14, 8, and 3 mg/L respectively. In vitro aminoacylation assays were performed with tritiumlabeled leucine and either E. coli or Halobacterium NRC-1 total tRNA (Figure 3, Panels A and B). Based on the charging of 10 µM crude total tRNA, MtLRS and AfLRS charge halobacterial tRNA 54- and 21- fold more efficiently than E. coli tRNA, respectively. The MjLRS enzyme, however, shows only a 6-fold preference for halobacterial tRNA. The E. coli enzyme was 100-fold more efficient at charging E. coli crude total tRNA than halobacterial tRNA. Therefore, MtLRS and AfLRS are good candidates for orthogonal aminoacyl-tRNA synthetases with respect to E. coli tRNA, but MjLRS is not. Since MtLRS showed a higher level of suppression with HL(TAG)1 in vivo than did AfLRS, the

Tilling .

....

MtLRS/HL(TAG)1 pair was carried forward as a potential new orthogonal pair for use in *E. coli*.

Optimization of the tRNA anticodon loop. The robust endogenous amber [0193] suppressor supD confers survival to 1000 µg/mL ampicillin when expressed from pACKO--A184TAG. In contrast, cells expressing the MtLRS/HL(TAG)1 pair survive to only 35 µg/mL ampicillin, which corresponds to a 2.9% suppression efficiency as determined from β -galactosidase assays (Table 1). We therefore sought to improve the activity of the system. Previous experiments on frameshift, missense, and nonsense suppression revealed that A37 was a highly conserved feature in robust suppressor tRNAs (Magliery et al., (2001) J. Mol. Biol. 307, 755-769). HhL4 has a G at position 37, therefore substitution of 10 G37 to A might be expected to improve suppression efficiency. To examine this and other possible anticodon loop mutants, a library was constructed in which the 7 positions of the anticodon loop (positions 32-38, see Figure 4, Panel A) in HhL4 were replaced with degenerate bases and subcloned into pACKO-A184TAG. The library of tRNAs was cotransformed with pKQ-MtLRS and subjected to ampicillin selection initially at 35 µg/mL 15 ampicillin for two rounds of selection, then plated on a series of plates with increasing ampicillin concentration in the third round of selection. At the highest concentration of ampicillin for which growth was observed (500 µg/mL), the only clone found had an anticodon loop with the sequence CUCUAAA, corresponding to a simple G37A mutation (Table 1). When cotransformed with pKQ-MtLRS, this clone could survive to $500~\mu g/mL$ 20 ampicillin. In the absence of the synthetase it survived to only 25 µg/mL ampicillin. Under similar conditions, cells containing the wild-type M. jannaschii tyrosyl orthogonal amber suppressor tRNA survive to 350 µg/mL ampicillin in the presence of the cognate synthetase and to 60 ug/mL ampicillin without the synthetase.

Table 1. Suppression efficiency of mutant orthogonal tRNAs.

Reporter Plasmid	Miller U	nits	
pLASC-lacZ(Leu)	210	±	2
pLASC-lacZ(Ser)	200	±	5
pLASC-lacZ(Tyr)	192	- ±	7
pLASC-lacZ(TAG)	1	±	1
pLASC-lacZ(AGGA)	2	±	1
pLASC-lacZ(TGA)	1	±	1

	Percent Suppression*						1		
Suppressor tRNA	with pKQ			with synthetase		ase	Sequence		
HL(TAG)1	0.4	±	0.1%	2.9	±	0.8%	GCGAGGGTAGCCAAGCTCGGCCAACGGCGACGGACTCTAGATCCGTTCTCGTAGGA GTTCGAGGGTTCGAATCCCTTCCCT		
HL(TAG)2	0.3	±	0.1%	9.6	±	0.4%	GCGAGGGTAGCCAAGCTCGGCCAACGGCGACTCTAAATCCGTTCTCGTAGGAGTTCGAAGGGTTCGAATCCCTTCCCTCGCACCA		
HL(TAG)3	1.5	±	1.2%	33.2	±	4.4%	CCCAGGGTAGCCAAGCTCGGCCAACGGCGACTCTAAATCCGTTCTCGTAGGAGTTCGAGGGTTCGAATCCCTTCCCTGGGACCA		
HL(AGGA)1	0.4	±	0.1%	4.6	±	2.1%	GCGAGGGTAGCCAAGCTCGGCCAACGGCGGACTTCCTAATCCGTTCTCGTAGG AGTTCGAGGGTTCGAATCCCTTCCCT		
HL(AGGA)2	0.7	±	0.3%	14.9	±	6.1%	GCGAGGGTAGCCAAGCTCGGCCAACGGCGACGCACTTCCTAATCCGTTCTCGTAGGAGTTCGAGGGTTCGAATCCCTCCC		
HL(AGGA)3	7.4	±	0.4%	35.5	±	1.4%	GCGGGGGTTGCCGAGCCTGGCCAAAGGCGCCGGACTTCCTAATCCGGTCCCGTAGG GGTTCCGGGGTTCAAATCCCGGCCCCGCACCA		
HL(TGA)1	4.7	±	1.5%	60.8	±	7.0%	GCGGGGGTTGCCGAGCCTGGCCAAAGGCGCCGGACTTCAAATCCGGTCCCGTAGGGGTTCCGGTACCCGCCCCGCACCA		
J17 ^b	0.2	±	0.1%	18.5	±	4.8%	CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCGCATGGCGCTGGTT CAAATCCGGCCCGCCGGACCA		
SupD	42.8	±	7.1%	ND			GGAGAGATGCCGGAGCGGCTGAACGGACCGGTCTCTAAAACCGGAGTAGGGGCAAC TCTACCGGGGGTTCAAATCCCCCTCTCTCCGCCA		
Ser2AGGA	25.2	±	0.1%	ND			GGAGAGATGCCGGAGCGGCTGAACGGACCGGTCTTCCTAAACCGGAGTAGGGGCAA CTCTACCGGGGGTTCAAATCCCCCTCTCCCGCCA		

5

10

"β-Galactosidase activity was determined for tRNA reporter plasmids derived from pACKO-Bla cotransformed with the appropriate pLASC-lacZ mutant and either a synthetase-expressing plasmid or a plasmid with no synthetase. Activity is reported as the percentage of activity observed relative to the value observed from the pLASC-lacZ construct with a leucyl (wild-type), seryl, or tyrosyl sense codon at position 25. In each case, the codon at position 25 of *lacZ* is designated in parentheses.

^bJ17, the *M. jannaschii* tyrosyl amber suppressor tRNA with improved orthogonality (Wang and Schultz (2001) <u>Chem. Biol.</u> 8, 883-890) was expressed in plasmid pACKO-A184TAG in the presence of pLASC-lacZ(TAG) and either pKQ or pBK-JYRS.

- 15 [0194] Randomization of leucyl acceptor stem. Although the activity of the HhL4-derived amber tRNA was significantly improved with the G37A mutation, the suppression level in the absence of the synthetase increased from 5 to 25 μg/mL ampicillin. To overcome the undesired increase in background suppression, a mutant of the HhL4-derived tRNA was sought that would not cross react with E. coli aminoacyl-tRNA synthetases.
- Almost all of *E. coli* synthetases recognize bases within the acceptor stem of their cognate tRNAs. Therefore, we anticipated that mutations within this region of the tRNA might eliminate interactions between the orthogonal tRNA and the cross-reactive synthetase. A

.....

library in which the 3 terminal base pairs of the acceptor stem and the discriminator base were randomized (positions 1-3 and 70-73, the randomized region is outlined in **Figure 4**, **Panel A**) was constructed from the HL(TAG)2 mutant tRNA and subcloned into pACKO-A184TAG.

- [0195] To identify members of this tRNA library that retained activity but were even poorer substrates for endogenous synthetases, a selection strategy was adopted from previous work on the M. jannaschii system (Wang and Schultz (2001) Chem. Biol. 8, 883-890). To isolate a pool of mutant tRNAs that had comparable activity to the G37A mutant of the HhL4-derived tRNA, the tRNA library in which the acceptor stem was randomized 10 was cotransformed with pKQ-MtLRS and subjected to two rounds of positive selection at 500 µg/mL ampicillin. Six clones surviving the positive selection were sequenced, and all were unique and conserved the discriminator base, A73 (Figure 4, Panels A and B). In all cases the stem positions had standard Watson-Crick base pairs. To identify members of the pool of active clones that would not be charged by endogenous aminoacyl-tRNA 15 synthetases, the surviving tRNA-expressing plasmids were transferred into cells containing a barnase reporter plasmid, pSCB2. This plasmid contains the gene for the RNase, barnase, with two TAG codons at permissive positions 2 and 44, under control of the arabinose promoter, as well as the gene for β -lactamase. Any tRNA that is aminoacylated by an endogenous E. coli synthetase will result in suppression of the nonsense codons and cell 20 death. The cells were plated on LB plates containing 25 µg/mL of chloramphenicol, 50 µg/mL of ampicillin to maintain the plasmids, and 0.2% arabinose to induce expression of the barnase gene. Sixteen survivors were sequenced, and three unique sequences were identified. All three clones had reversed the 3:70 base pair from G:C to C:G. Of these, mutant HL(TAG)3 gave the highest level of suppression in the presence of MtLRS 25 (600 μg/mL ampicillin) and only survived to 7.5 μg/mL ampicillin without the synthetase. These values correspond to 33.2% suppression in the presence of MtLRS and 1.5% in the absence of the synthetase as determined by β-galactosidase amber suppression assays (see Table 1). By comparison, the mutant M. januaschii suppressor tRNA, J17, gives values of 18.5% and 0.2% with and without the *M. jannaschii* tyrosine synthetase, respectively.
- 30 [0196] Identification of AGGA suppressors. To expand the list of codons that can be used for unnatural amino acid mutagenesis, a tRNA that could efficiently suppress a four-base codon was sought. Previous studies indicated that the four-base codon AGGA

can be efficiently suppressed in *E. coli*, and tRNAs with 8 nucleotide anticodon loops were the most efficient suppressors of AGGA codons (Magliery et al., (2001) J. Mol. Biol. 307, 755-769). A β-lactamase reporter plasmid analogous to the TAG reporter was constructed but with A184 replaced by AGGA instead of TAG. Normal translation in the absence of a +1 frameshift suppressor tRNA should result in missense errors downstream of position 184 and premature truncation of the protein. A library of tRNAs derived from the HhL4 tRNA was constructed in which the 7 nucleotide anticodon loop was replaced with 8 random nucleotides. The library was subcloned into pACKO-A184AGGA, cotransformed with pKQ-MtLRS, and then subjected to ampicillin selection. At the highest concentration of ampicillin at which growth was observed, 75 μg/mL, only one clone, HL(AGGA)1, was found. This clone had the anticodon loop sequence CUUCCUAA. As was the case with the *bla* A184TAG reporter plasmid, cells transformed with pACKO-A184AGGA can survive to only 5 μg/mL ampicillin in the absence of a suppressor tRNA. Therefore, the clone identified, HL(AGGA)1, is a weak suppressor of AGGA codons.

10

30

15 [0197]During these experiments, serendipitous mutants capable of surviving up to 300 µg/mL ampicillin were identified. These mutants were no longer orthogonal, and all had multiple point mutations relative to the parent sequence. All of the clones contained the substitution T65C. This mutation corrects the G:U mismatch present in the TψC loop stem, suggesting that this G:U base pair might be detrimental to suppressor activity. We therefore 20 decided to randomize this base; a library was made in which the 49:65 base pair was randomized in HL(AGGA)1. The library was subcloned into pACKO-A184AGGA, and then cotransformed with pKQ-MtLRS. Of the 16 library members, the most efficient suppressor, HL(AGGA)2, was identified by ampicillin selection. This clone contained a T65C mutation and could survive to 125 µg/mL ampicillin. Nevertheless, this level of 25 activity was far lower than that observed for the corresponding amber suppressors. Consequently, alternative strategies were considered.

[0193] Mutations in the D-loop have been previously implicated in frameshift suppression (Tuohy et al., (1992) <u>J. Mol. Biol.</u> 228, 1042-1054), and we next hypothesized that such mutations might improve the suppression efficiency of the AGGA suppressor. Libraries wherein the 13 nucleotides of the D-loop (position 14-21, see **Figure 4, Panel A**) were replaced with 11 or 13 random nucleotides were prepared in pACKO-A184AGGA.

Although a great deal of sequence diversity was observed among the survivors at the

highest concentrations of ampicillin (125 μ g/mL), no mutants were observed with increased activity relative to the parent tRNA.

[0199] A consensus-derived AGGA suppressor tRNA. In examining the sequence of the HhL4-derived tRNA, there was no obvious explanation for the poor activity of this 5 - suppressor. Rather than mutate HL(AGGA)2 further, we pursued an alternative approach. The archaeal leucyl-tRNAs are highly similar, varying from each other usually by only a few base substitutions (Figure 1, Panel C). The entire family would be well represented by a library derived from a consensus sequence with many random mutations throughout. The consensus sequence was compiled with the GCG program pileup, and those positions 10 considered degenerate by the program were changed to the most frequent base at those positions. The anticodon loop was changed to CUUCCUAA since this sequence was already shown to be the optimal sequence for an AGGA suppressor derived from HhL4. The final sequence used as the consensus sequence is shown in Figure 5. A library was synthesized by overlap extension of oligonucleotides in which each site of the tRNA gene 15 was synthesized as a doped mixture of 90% the consensus sequence and 10% a mixture of the other 3 bases. The library was subcloned into pACKO-A184AGGA. Sequencing of 24 naïve clones revealed that the average number of mutations per clone was 5.9, and these mutations were randomly distributed throughout the tRNA sequence. After cotransformation with pKQ-MtLRS and selection on ampicillin plates, several clones 20 survived to 300 µg/mL of ampicillin and were found to be the original sequence with the 27:42 and 49:65 base pairs changed to the canonical base pairs T27:A42, G27:C42, or C27:G42, and G47:C65 or C47:G65 (Figure 5). The most efficient suppressor, designated HL(AGGA)3, can survive to 300 µg/mL ampicillin in the presence of pKQ-MtLRS but to only 30 µg/mL in the absence of the synthetase, which correspond to 35.5% and 7.4% 25 suppression, respectively, as determined by β -galactosidase assays (Table 1).

[0200] Identification of opal suppressor tRNAs. To further expand the list of codons, we sought opal suppressors derived from HhL4. A reporter plasmid, pACKO-A184TGA, was constructed in which the A184 position of β -lactamase was changed to TGA. This bla A184TGA reporter plasmid can survive to 10 μ g/mL ampicillin without any suppressor tRNA present, whereas the TAG and AGGA reporters could survive to only 5 μ g/mL. In the case of opal suppression, there is background read-through that leads to the

30

production of a small amount of protein even in the absence of a suppression system. Nevertheless, this level is quite small. To identify suppressors, a library in which the anticodon loop (positions 31-38) of HhLA was replaced with 7 degenerate nucleotides was prepared in pACKO-A184TGA. When cotransformed with pKQ-MtLRS, no members of this library could survive on ampicillin plates at 50 μg/mL. Instead of HhLA, a library was prepared in which the 8 nucleotide anticodon loop was randomized with 7 nucleotides in HL(AGGA)3, the most robust AGGA suppressor identified from the consensus sequence. At the highest concentrations of ampicillin at which growth was observed (300 μg/mL) only one clone, designated HL(TGA)1, with the sequence CUUCAAA was found. The clone can survive to 350 μg/mL ampicillin when coexpressed with pKQ-MtLRS, but can survive to only 30 μg/mL without the synthetase plasmid, which corresponds to 60.8% suppression as determined by β-galactosidase assays (Table 1). Apparently, the beneficial effects of using the consensus sequence are not limited to frameshift suppression.

10

15

20

25

30

orthogonal tRNA-synthetase pairs is to adapt eukaryotic or archaeal synthetases and tRNAs for use in *E. coli*. Several yeast synthetases, notably glutamine, aspartic acid, arginine, and tyrosine, have been shown not to recognize *E. coli* tRNA, and might therefore be useful for the construction of orthogonal tRNA-synthetase pairs. Unfortunately, many eukaryotic synthetases express poorly or have low specific activity in *E. coli*. Eukaryotic synthetases, particularly the mammalian enzymes, are often organized into large complexes (Mirande et al., (1982) EMBO J. 1, 733-736), and the low activity often observed may be related to the inability to form these complexes in *E. coli*.

[0202] The success of the *M. jannaschii* tyrosyl orthogonal pair (Wang et al., (2001) Science 292, 498-500) suggested that archaebacteria may in general be a good source of orthogonal pairs. Early work on the halophile *Halobacterium cutirebrum* (Kwok and Wong (1980) Can. J. Biochem. 58, 213-218) indicated that almost all the tRNAs of this archaean (notably leucine, arginine, tyrosine, serine, histidine, and proline) cannot be charged by *E. coli* aminoacyl-tRNA synthetases. Indeed, archaeal tRNA synthetases are more similar to their eukaryotic than prokaryotic counterparts in terms of homology and tRNA recognition elements. Unlike their eukaryotic counterparts, however, there is currently no evidence for their higher order assembly into structured multimers (Tumbula et al., (1999) Genetics 152, 1269-1276; Woese et al., (2000) Microbiol. Mol. Biol. Rev. 54, 202-236). Moreover, since

14.

10

15

20

25

30

most archaea are thermophiles, active synthetases from archaea can be expressed in good yields in *E. coli* and can be readily purified in active form. Due to extensive sequencing efforts, at least 16 archaeal genome sequences are currently available, which together with the lack of introns in the genome, greatly facilitates the PCR amplification of the archaeal synthetase genes. For all of the above reasons, our attention has focused on the archaea as a source for orthogonal pairs.

[0203] Another design issue in the construction of orthogonal tRNA-synthetase pairs is the ability of the aminoacyl-tRNA synthetase to recognize mutants of the cognate tRNA with altered anticodon loops (i.e., nonsense or missense suppressors). AminoacyltRNA synthetases frequently use the anticodon loop as a major positive identity element, and mutations in this region of the tRNA frequently result in impaired synthetase recognition. The leucyl-tRNA synthetases frequently lack strong anticodon recognition elements, and a leucyl orthogonal tRNA-synthetase pair can therefore be able to decode a variety of codons, including amber, opal and four-base codons. Of the archaeal leucyltRNA synthetases, only the enzyme from Haloferax volcanii has been thoroughly investigated (Soma et al., (1999) J. Mol. Biol. 293, 1029-1038). The synthetase does not recognize bases in the anticodon loop; instead, a highly conserved pattern of mismatches within the variable loop is the primary recognition element for the synthetase. Although the cloning of the gene for this enzyme has not been reported, the sequenced genome of a closely related archaean, Halobacterium sp. NRC-1, is available. A multiple sequence alignment of leucyl-tRNA synthetases from many phyla including archaeal, prokaryotic, and eukaryotic sequences (Figure 1, Panel A) shows that the halophilic enzyme is unusual among the family of archaeal synthetases, having greater homology to the prokaryotic branch than the eukaryotic or archaeal branches. Unlike the synthetases, all archaeal leucyl tRNAs are highly homologous and share absolutely conserved features such as A73, G37, and a 12 nucleotide variable loop with 2 unpaired bases (Figure 1, Panel B). The conservation of these positive recognition elements led us to believe that tRNA recognition by the other archaeal leucyl-tRNAs would be similar to recognition by the halobacterial synthetase. Consequently, these synthetases can be useful in the construction of orthogonal tRNA-synthetase pairs when combined with suppressor tRNAs derived from archaeal leucyl-tRNAs.

5

10

15

20

25

30

[0204] Because archaeal leucyl-tRNAs and synthetases are highly homologous to one another, both species-specific and cross-species combinations could potentially function as efficient orthogonal tRNA-synthetase pairs. Therefore, each of the five potential orthogonal tRNAs (AfL3, HhL4, MjL2, PfL5, and PhL2) was examined in the presence of each of the six separate archaeal synthetases (AfLRS, ApLRS, HhLRS, MjLRS, MtLRS, and PhLRS) for the ability to suppress A184TAG in bla. All five orthogonal tRNAs afforded a higher level of amber suppression in the absence of an archaeal synthetase than is observed when no amber suppressor tRNA is present in the cell. All five suppressors are, therefore, expressed, processed, and functionally charged to some degree by an endogenous E. coli synthetase. Nevertheless, only three of the five tRNAs (PhL2, AfL3, and HhL4) gave a higher level of suppression when a foreign synthetase (either MjLRS, MtLRS, or AfLRS) was coexpressed with the tRNA than was observed with no synthetase. The MjL2 and PfL5 suppressors fail to give an enhancement in suppression when coexpressed with a cognate or noncognate archaeal synthetase. Without being limited to one theory, these tRNAs may be expressed as functional suppressor tRNAs in E. coli but are unable to be charged due to incompatibility with both cognate and noncognate synthetases. In the case of MjL2, the suppressor is derived from the natural substrate for MjLRS, so it seems unlikely that the tRNA would not be charged, when other tRNAs are efficiently charged by MjLRS. Another explanation might be that the tRNAs are efficiently charged but are incompatible with the E. coli translational machinery, but this is not consistent with the fact that some suppression is observed when no archaeal synthetase is present. Another possibility is that MjL2 and PfL5 are efficiently charged with leucine, but are deacylated in an editing process by an endogenous E. coli synthetase. In any case, not all archaeal leucyl isoacceptors are equivalent in their ability to function as orthogonal amber suppressors in E. coli.

Only three of the six leucyl-tRNA synthetases (MjLRS, MtLRS, and AfLRS) cloned from archaea gave a higher level of suppression when combined with any of the five orthogonal tRNAs. In the case of HhLRS, the synthetase does not yield protein when over-expressed. Without being limited to one theory, it is most likely, PhLRS and ApLRS do not express functional protein in *E. coli* either, but it is also possible that the proteins are not active at 37°C, or do not recognize any of the orthogonal tRNAs tested. There was no evidence that some tRNAs are preferred substrates for a specific synthetase. Indeed,

 $u_{i} + u_{i}$

20

25

30

although a tRNA from *M. jannaschii* was one of the five orthogonal tRNAs examined, the halobacterium-derived suppressor was the preferred substrate for MjLRS. All three functional tRNAs gave the highest level of suppression when charged by MtLRS or MjLRS, and to a lesser degree with AfLRS.

5 -[0206] Although on the whole the archaeal leucyl synthetases have similar tRNA recognition properties, it is clear from *in vitro* charging experiments that there are some differences in their recognition of tRNA. The charging of crude total *E. coli* tRNA by AfLRS and MtLRS is only 5- and 13-fold higher, respectively, than the background reaction observed with no synthetase, whereas MjLRS is able to charge *E. coli* tRNA 50-fold over background. Such differences in tRNA recognition among highly homologous synthetases was unanticipated, but not without precedent (Kwok and Wong (1980) Can. J. Biochem. 58, 213-218). Since aminoacyl-tRNA synthetases have evolved only to be orthogonal to the non-cognate tRNAs present in their own host's cytoplasm, it is perhaps not surprising that subtle variations in sequence or chemical modification can lead to mischarging in foreign systems.

Improving the activity of orthogonal suppressor tRNAs. To date, we have [0207] identified and characterized three orthogonal tRNA-synthetase pairs: the yeast glutamine (Liu and Schultz (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 4780-4785), yeast aspartate (Pastrnak et al., (2000) Helv. Chim. Acta 83, 2277-2286), and archaeal tyrosine pairs (Wang et al., (2000) J. Am. Chem. Soc. 122, 5010-5011). Of these systems, only the tyrosine system gives levels of amber suppression comparable to the levels observed for strong native amber suppressors such as supD or supF. When expressed with a the high-copy βlactamase reporter pBLAM (the reporter plasmid for this study was a medium-copy plasmid) in the presence of their cognate synthetase, cells containing the original glutamine, aspartate, and tyrosine orthogonal amber suppressor tRNAs can survive to 140, 60, and 1220 ug/mL ampicillin, respectively (Pastrnak et al., (2000) Helv. Chim. Acta 83, 2277-2286; Wang et al., (2000) J. Am. Chem. Soc. 122, 5010-5011). A high level of suppression may be critical to the successful modification of the amino acid specificity of synthetases using a double-sieve selection strategy (Liu and Schultz (1999) Proc. Natl. Acad. Sci. U. S. A. 96, 4780-4785). For suppression systems with low activity, it is often difficult to distinguish active and inactive synthetases in selection experiments due to their similarity in

phenotype. A high level of suppression is required for the production of protein containing

unnatural amino acids. Therefore, a great deal of attention has been paid to those features of orthogonal tRNAs that give rise to robust suppression.

[0208] Previous work on frameshift and amber suppression in E. coli clearly indicates that positions 31, 32, 37, and 38 of the tRNA anticodon loop have profound effects on -suppression efficiency (Yarus et al., (1986) J. Mol. Biol. 192, 235-255; Smith et al., (1987) Nucleic Acids Res. 15, 4669-4686; Raftery and Yarus (1987) EMBO J. 6, 1499-1506; Kleina (1990) J. Mol. Biol. 213, 705-717). The presence of G37 in all the archaeal leucyl tRNAs led us to believe that a substitution at this position might lead to a higher suppression efficiency. Indeed, randomization of the anticodon loop showed that the most 10 efficient suppressors have the anticodon loop CUCUAAA. Although the tRNA was toxic, the G37A mutant also emerged through selection with the M. jannaschii tyrosine system (Wang and Schultz (2001) Chem. Biol. 8, 883-890) as the most potent suppressor thus far observed for this system. Similar selection experiments with the yeast-derived glutamine and aspartate orthogonal pairs have been performed in which libraries of positions 32-38 of the anticodon loop are replaced with degenerate bases then subjected to positive ampicillin 15 selection in the presence of the cognate synthetase (J.C.A., P.G.S., and Miro Pastrnak, unpublished results). In both cases, the anticodon loop sequence CUCUAAA afforded the highest suppression efficiency corresponding to six-fold and five-fold enhancements in the concentration of ampicillin at which growth is observed for the glutamine and aspartate systems, respectively. In at least three other systems, tRNAs with the anticodon loop 20 sequence CUCUAAA have emerged as the most efficient amber suppressors. The anticodon loop sequence CUUCCUAA was found to be the most efficient sequence for a leucyl AGGA suppressor. Selection experiments on tRNAs with randomized anticodon loops in E. coli tRNA ser similarly converged on the anticodon loop sequence CUUCCUAA for AGGA suppression (Magliery et al., (2001) J. Mol. Biol. 307, 755-769), and the sequence CUUCAAA also emerged as the most efficient anticodon loop sequence for a leucyl opal suppressor. These results suggest that the preferred anticodon loop sequence is determined by interactions with endogenous translational machinery rather than the particular preferences of the aminoacyl-tRNA synthetases. Indeed, the anticodon loop may 30 require sequence-specific modifications in order to function optimally (Soderberg and Poulter (2000) Biochemistry 39, 6546-6553; Sussman and Kim (1976) Science 192, 853-858). Alternatively, Yarus (Yarus (1982) Science 218, 646-652) has suggested that the

entire anticodon stem and loop (positions 27-43 of the tRNA) together function as an "extended anticodon" that interacts with ribosome as a module. The entire sequence of this region can help to define the identity of the anticodon for proper decoding.

tRNAs with the sequence CU(X)XXXAA in the anticodon loop. Although this consensus sequence is preferred for TAG, TGA, and AGGA codons, other sequences may be preferable for other four- and five-base codons. In previous studies (Magliery et al., (2001) J. Mol. Biol. 307, 755-769), the most efficient suppressor tRNAs had bases at positions 32, 33, 37, and 38 which differed from the consensus sequence. For example, the most efficient suppressors of the codon CUAG had an anticodon loop with the sequence CGCTAGGA, deviating at both U33 and A37. In addition, some synthetases employ position 37 as a strong positive determinant for recognition, in which case a CU(X)XXXAA anticodon loop sequence can prove to be non-optimal.

10

[0210] Optimization of the anticodon loop sequence as described above was 15 sufficient to provide an efficient amber suppressor tRNA for the leucine system. Optimization of the anticodon loop of the AGGA frameshift suppressors derived from HhLA also afforded a viable tRNA. However the suppression efficiency (4.6%) of this tRNA, HL(AGGA)1, is far lower than that measured for the suppression of amber codons by HL(TAG)2. Indeed, this suppressor permitted survival at only 75µg/mL ampicillin, 20 significantly less than the seryl AGGA suppressor (Ser2AGGA) identified previously (Magliery et al., (2001) J. Mol. Biol. 307, 755-769), which can survive to 275 ug/mL ampicillin when expressed in plasmid pACKO-A184AGGA. In general, the best AGGA suppressors are less active than the best amber suppressors (Anderson et al., (2002) Chem. Biol. 9, 237-244), but there appears to be something particular to HhL4 that hinders its 25 ability to act as a frameshift suppressor. The only feature obviously different from robust four-base suppressors previously identified (Atkins et al., (1991) Annu. Rev. Genet. 25, 201-228) is the presence of a very large D loop in HhL4. Most suppressors have 9 nucleotides in the D loop and 4 base pairs in the stem. HhL4 has only 3 base pairs in the stem and 13 bases in the loop. Moreover, previous studies have shown the D loop to play a 30 role in frameshift suppression (Tuohy et al., (1992) J. Mol. Biol. 228, 1042-1054). Not only did we see no increase in activity upon randomization of the D loop, there was also a great deal of sequence variation among the most active suppressors.

[0211] The serendipitous appearance of mutations in the G49:U65 base pair of the fourbase suppressor tRNAs suggested that non-canonical base pairing in the stem regions of tRNAs has a deleterious effect on suppression efficiency. This hypothesis was further supported by a randomization and selection experiment on the acceptor stem of the HhL4derived amber suppressor. The three terminal base pairs of the acceptor stem were simultaneously randomized. This library of tRNAs would therefore contain all combinations of mismatched and Watson-Crick base pairs. In fact, 98.4% of the theoretical members of this library should have at least one mismatched base pair. Nevertheless, in the 9 active acceptor stem mutants outlined in Figure 4, Panel B, all positions are occupied by Watson-Crick base pairs. Similarly, the D, TψC, anticodon, and acceptor stems of the yeast glutamine amber suppressor tRNA have been individually randomized and subjected to positive selection (J.C.A. and P.G.S., unpublished results). In all surviving clones, every position in these stem regions was occupied by a Watson-Crick pair. In the parent tRNA, the 6:67 base pair is U:G. Mutation of this base pair to U:A results in a doubling of the concentration of ampicillin at which cells can grow. Also, when subjected to positive selection, the only mutations that emerged from random mutagenesis of the leucyl consensus-derived frameshift suppressor appeared at mispaired sites. Others have also noted that mispairing in stem regions adversely affects suppression efficiency (Buttcher et al., (1994) Biochem. Biophys. Res. Commun. 200, 370-377; Hou et al., (1992) Biochemistry 31, 4157-4160). Without being limited to one theory, it may be that tRNAs with mispaired bases are not readily folded into the correct cloverleaf structure and therefore are not readily processed and modified (Furdon et al., (1983) Nucleic Acids Res. 11, 1491-1505). A quantitative analysis of the ratio of charged to uncharged species and of the ratio of fully processed to unprocessed tRNA present in the cell could enhance our understanding of the mechanisms by which these poorly-suppressing tRNAs are impaired. [0212]An analysis of multiple sequence alignments of many families of tRNAs reveal multiple examples of conserved non-Watson-Crick pairings. For example, a G3:U70 base pair is a conserved positive determinant for recognition by E. coli alanyl-tRNA synthetase (Martinis and Schimmel (1995) in tRNA: Structure, Biosynthesis, and Function (Soll, D., and RajBhandary, U., Eds.) pp 349-370, ASM Press, Washington, DC.). If the element is a conserved positive determinant for recognition, then it may prove difficult to construct robust suppressor tRNAs for the cognate synthetase. Most often, however, the

10

15

20

25

30

mispairing present in native sequences is only found in specific isoacceptors. Without being limited to one theory, these and other variations from the consensus sequence of the family of tRNAs present in individual isoacceptors may be present as a result of subtle, species-specific adaptations in positive or negative synthetase recognition, optimal processing and modification, or interactions with elongation factors. Alternatively, these variations may simply be the result of neutral evolutionary drift.

[0213] When transferred to another species, these variations are unlikely to offer to the new host's translational machinery any advantages they conferred to the source organism. Furthermore, for cross-species pairs, the synthetase is unlikely to recognize any species-specific identity elements present in the tRNA. Only those recognition elements common to the entire family are likely to be useful. Similarly, any processing or modification adaptations particular to a specific tRNA would be lost to the *E. coli* translational apparatus. These variations may even be deleterious to suppression efficiency, particularly when these variations are mismatched bases in stem regions. Suppressors tRNAs derived from the consensus sequence preserve only those features that are broadly shared by the entire family, and eliminate potentially deleterious variations. Therefore, suppressor tRNAs derived from the consensus sequence may in general lead to higher suppression efficiencies.

10

15

20

25

30

[0214] Although optimization of the anticodon loop and elimination of mispairing gave modest to large increases in suppression efficiency, these modifications were not sufficient to provide robust AGGA and opal suppressor tRNAs. Only the consensus-derived suppressors had activities comparable to the tRNA₂^{ser}-derived suppressors described previously (Anderson et al., (2002) Chem. Biol. 9, 237-244). A comparison of the consensus-derived sequences for HL(AGGA)3 and HL(AGGA)2 reveal that there are 14 base substitutions, but neither sequence has mispairs. Without being limited to one theory, perhaps by using the consensus sequence of the entire family of tRNAs, those bases that are specific to any particular tRNA and may be detrimental to activity are identified and eliminated.

[0215] Improving the selectivity of orthogonal suppressor tRNAs. Unfortunately, improvements in the activity of these suppressor tRNAs also brought about an undesirable increase in the level of suppression observed in the absence of synthetase. The original M.

jannaschii tyrosine orthogonal suppressor tRNA was partially charged by an *E. coli* synthetase, but the reaction was eliminated by mutagenesis (Wang and Schultz (2001) Chem. Biol. 8, 883-890). A double sieve selection was able to identify mutants of the wild-type tRNA with excellent orthogonality, but there was also a significant loss of overall activity. Ideally, mutations could be introduced into the tRNA that would eliminate the cross-reactivity with *E. coli* synthetases but preserve high levels of suppression efficiency. Since aminoacyl-tRNA synthetases frequently recognize positions within the acceptor stem and discriminator base of tRNAs (Martinis and Schimmel (1995) in tRNA: Structure, Biosynthesis, and Function (Soll and RajBhandary, Eds.) pp 349-370, ASM Press, Washington, DC.), it is likely that an *E. coli* synthetase that charges the orthogonal tRNA would have a positive recognition element in this region. If this determinant could be changed without destroying recognition by the foreign synthetase, activity could be preserved while eliminating the background reaction. When this strategy was applied to the HhL4-derived amber suppressor, such mutants were indeed found. Several mutants preserved or even improved suppression efficiency when coexpressed with MtLRS but had

10

15

20

25

30

preserved or even improved suppression efficiency when coexpressed with MtLRS but had nearly background levels of amber suppression (7.5 versus 5 ug/mL ampicillin) in the absence of the synthetase. These mutants had reversed the third base pair from G:C to C:G, and an inspection of the recognition elements known for various *E. coli* synthetases suggests the identity of the *E. coli* synthetase that had cross-reacted with HhLA-derived suppressor. Both GlnRS and LysRS of *E. coli* conserve G3:C70 and frequently cross-react with amber suppressor tRNAs (Kleina (1990) J. Mol. Biol. 213, 705-717). Because LysRS also conserves A73, this is the more likely candidate for the cross-reactive *E. coli* synthetase (Freist and Gauss (1995) Biol. Chem. Hoppe-Seyler 376, 451-472; McClain et al., (1988) Science 242, 1681-1684). This strategy can be a general solution to the problem of improving the specificity of cross-reactive orthogonal tRNAs since most *E. coli* aminoacyl-

[0216] We have shown that the leucyl-tRNA synthetase from the archaean Methanobacterium thermoautotrophicum and mutants of a halobacterial tRNA function as an orthogonal pair in E. coli. Mutagenesis experiments showed that the two most significant criteria that lead to efficient orthogonal amber suppressor tRNAs are a CU(X)XXXAA anticodon loop and the lack of non-canonical or mismatched base pairs in the stem regions. From these selections, we have identified efficient amber, four-base, and opal orthogonal suppressor tRNAs. We have also devised a consensus strategy to rationally

tRNA synthetases contain positive determinants within the acceptor stem.

design efficient orthogonal tRNAs. This leucyl-orthogonal pair can be combined with the *M. jannaschii* pair to site-specifically incorporate two unique unnatural amino acids simultaneously into proteins *in vivo*.

[0217] Abbreviations and Textual Footnotes: Af, Ap, Hh, Mj, Mt, Pf, Ph, and Ec:
-Archaeoglobus fulgidus, Aeuropyrum pernix, Halobacterium sp. NRC-1, Methanococcus jannaschii, Methanobacterium thermoautotrophicum, Pyrococcus furiosus, Pyrococcus horikoshi, and Escherichia coli, respectively; LRS, leucyl-tRNA synthetase; bla, gene for β-lactamase; lacZ, gene for β-galactosidase.

EXAMPLE 2: EXEMPLARY LEUCYL O-RSs AND LEUCYL O-tRNAs.

10 [0218] Exemplary O-tRNAs comprise, e.g., SEQ ID NO.:1-7 and 12 (See, Table 3). Exemplary O-RSs include, e.g., SEQ ID NOs.: 15 and 16 (See, Table 3). Exemplary polynucleotides that encode O-RSs or portions thereof include, e.g., SEQ ID NOs.: 13 and 14.

[0219] Further details of the invention, and in particular experimental details, can be found in Anderson, John Christopher, "Pathway Engineering of the Expanding Genetic Code," Ph.D. Dissertation, The Scripps Research Institute [2003].

[0220] It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims.

20

25

[0221] While the foregoing invention has been described in some detail for purposes of clarity and understanding, it will be clear to one skilled in the art from a reading of this disclosure that various changes in form and detail can be made without departing from the true scope of the invention. For example, all the techniques and apparatus described above can be used in various combinations. All publications, patents, patent applications, and/or other documents cited in this application are incorporated by reference in their entirety for all purposes to the same extent as if each individual publication, patent, patent application, and/or other document were individually indicated to be incorporated by reference for all purposes.

TABLE 3

SEQ	Label	SEQUENCE
ון איני		· · · · · · · · · · · · · · · · · · ·
EQ ID: 1	HL (TAG) 1	GCGAGGGTAGCCAAGCTCGGCCAACGGCGACGGACT CTA GATCCG
,		
1		
SEQ ID. 2		
SEQ ID: 3	HL (TAG) 3	
SEQ ID. 3	tRNA	
SEQ ID: 4	HL (AGGA)	TOTAL ACCOUNT OF THE PROPERTY
SEQ ID: 4	1	GCGAGGGTAGCCAAGCTCGGCCAATCCCTTCCCTCGCACCA GTTCTCGTAGGAGTTCGAGGGTTCGAATCCCTTCCCT
İ	tRNA	TOTAL COCCOLOGIA CTTCCTAATCC
SEQ ID: 5	HL (AGGA)	GCGAGGGTAGCCAAGCTCGGCCAACGGCGACGGACTTCCTAATCC
52 4	2	GTTCTCGTAGGAGTTCGAGGGTTCGAATCCCTCCCTC
	tRNA	TO SECOND CONTROL A AGGCGCCGGACTTCCTAATCC
SEQ ID: 6	HL (AGGA)	GCGGGGTTGCCGAGCCTGGCCAAAGCCCCCCCCCCCCCC
	3	GGTCCCGTAGGGGTTCCGGGGTTCTZZZZZZ
	tRNA	GCGGGGGTTGCCGAGCCTGGCCAAAGGCGCCCGCCCCCGCACCA
SEQ ID: 7	HL (TGA) 1	The address of the state of the
	tRNA	GTCCCGTAGGGGTTCCGGGGTTCAAATCCGCA CCGGCGGTAGTTCAGCAGGGCAGAACGGCGGACTCTAAATCCGCA
SEQ ID: 8	J17	TGGCGCTGGTTCAAATCCGGCCCGGACCA
	M. jannaschii	TGGCGCTGGTTCAMTTGGGGG
	mtRNATyr	
	IIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIIII	
	 	GGAGAGATGCCGGAGCGGCTGAACGGACCGGTCTCTCAAAACCGGA
SEQ ID: 9	SupD	The same of the sa
	0.7003	
SEQ ID: 1	0 Ser2AGGA	GGAGAGATGCCGGAGCGGCTGAACGGTCCCCCTCTCTCCCGCC AGTAGGGGCAACTCTACCGGGGGTTCAAATCCCCCCTCTCTCCCGCC
	1 7 1 of	A GCGAGGGUAGCCAAGCUCGGCCAACGGCGACGCACCAAGAUCCG
SEQ ID:"I	l Leu4 of Halobact	
	erium	000000100110
	sp. NRC-	
	1	THE GOOD A DOCCOCCOGGACUUCCUAAUCC
SEQ ID:	12 Concensu	GCGGGGGUUGCCGAGCCUGGCCAAAGGCGCCGGACUUCCUAAUCC
	s-	GCGGGGGUUGCCGAGCCUMAGGGGUUCAAAUCCCCGCCCCCGCACCA
ŀ	derived	
	AGGA	
	Suppress	COMOCONCA A
	or	ATGAGCGATT TCAGGATAAT TGAGGAGAAG TGGCAGAAGG CGTGGGAGAA
SEQ ID:		ATGAGCGATT TCAGGATAAT TGAGGAGAAG TGGCACAGAAG TTTTTTCTCAAGGAGAAAT TTTGAGTCCG ATCCTAATGA GAAGGAGAAG TCACACGAGAAG TCACACGAGAATC TTCACGCAGG TCACACGAGAAATC
	lobus fulgidus	CAATTCCCTA TCCTTACCTT AATGGAAATC TOCATGAGAA TGAAGGGCTA
	leucyl	ACCTTCACAA TTGGCGATGC CTTCGCCAGA TACATCATTC CAACGTTCTC TTTCCCCTCG GCTTTCATGT TACGGGCACC CCAATCATTC
	tRNA-	CAACGTTCTC TTTCCCCTCG GCTTCATGT TACGGGACGAT AGAGGTTTAC GCCTTGCGGA GCTCATAGCC AAGAGGGACG AGAGGACGAT AGAGGTTTAC GCCTCGCACGAC TTGCTTCAGC TCACAACTCC
	syntheta	GCCTTGCGGA GCTCATAGCC AAGAGGGACG AGAGGGCTCAGC TCACAACTCC ACCAAATACC ATGACGTTCC GCTGGAGGAC TTGCTTCAGC TCACAACTCC
	\ -	ACCAAATACC ATGACGTTCC GCTGGAGGAC TIGCTTGAAGA AGAGAAAATC GTTGAGTACT TCTCAAGGGA GGCGCTGCAG GCTTTGAAGA GCATAGGCTA CTCCATTGAC TGGAGGAGGG TTTTCACCAC AACCGATGAA
1	lse	ACACCAMANIC CARACTEC CARACTEC CONTRACTOR CON

							
	ļ		GAGTATCAGA	GATTCATCG	A GTGGCAGTAC	TGGAAGCTCA	AGGAGCTTGG
	{	1	CCTGATTGTG	AAGGGCACC	C ACCCCGTCAG	ATACTGCCCC	CACGACCAGA
	ļ		ATCCTGTTGA	AGACCACGA	C CTTCTCGCTG	GGGAGGAGGC	' ልልርጥልጥጥርጥጥ
		1	GAATTTACCG	TTATAAAGT	r caggettgaa	GATGGAGACC	. TOTO TITLE OF T
		ļ	CTGTGCAACT	CTCCGTCCC	G AAACCGTGTT	TCCCCTCACA	· NACAMOMOCO
		ļ	TAAAGCCGAC	AACCTACGT	A ATTGCCGAGG	DOMOTED DE L	AACATCTGGG
			GTGAGCAAAG	ACCCUTACCI	A GAAGCTCACC	TGGATGGGGA	AAAGTGGTTT
			GCTGCTGGAC	Cyccumcym	COMOCOACO	TACACGGAGA	AAAAAGTCAG
	-		TOCCCCCTCCT	GAGGIIGAI	CGTCGCAGTT	CTTCGGCAAG	TACGTCATAG
			ACCCACAACC	AAACAGAAAA	GTGCCAATTC	TGCCTGCAGA	GTTTGTTGAC
			ACCGACAACG	CAACAGGAG	TGTGATGAGC	GTTCCCGCAC	ACGCTCCTTT
- 1			TGACCTGGCT	GCCATTGAGG	ACTTGAAGAG	AGACGAGGAA	ACGCTGGCGA
-			AGTACGGAAT	TGACAAAAGC	GTTGTAGAGA	GCATAAAGCC	AATAGTTCTG
-			ATTAAGACGG	ACATTGAAGG	TGTTCCTGCT	GAGAAGCTAA	TAAGAGAGCT
			TGGAGTGAAG	AGCCAGAAGG	ACAAGGAGCT	GCTGGATAAG	GCAACCAAGA
- 1			CCCTCTACAA	GAAGGAGTAC	CACACGGGAA	TCATGCTGGA	CAACACGATG
		1	AACTATGCTG	GAATGAAAGT	TTCTGAGGCG	AAGGAGAGAG	TTCATGAGGA
-			TTTGGTTAAG	CTTGGCTTGG	GGGATGTTTT	CTACGAGTTC	AGCGAGAAGC
- 1			CCGTAATCTG	CAGGTGCGGA	ACGAAGTGCG	ТТСТТААССТ	TCTTACCCAC
-			CAGTGGTTCC	TGAACTACTC	CAACAGAGAG	TGGAAGGAGA	ስርርምጥርጥር እ አ
-			TCACCTTGAA	AAGATGCGAA	TCATCCCCGA	CTACTACAAC	CACCACTTCA
1			GGAACAAGAT	TGAGTGGCTC	AGGGACAAGG	CINCIACAAG	AACCAAGTTCA
-1			CTTGGAACGA	CAATTCCCTC	GGATAAGGAG	TCCCTC A TCC	AAGGAAGGGG
1			AGACTCAACA	ATCTACATC	CCTACTACAT	TGGCTCATCG	AGAGCCTTTC
1			CAGGATTGCT	CYVCCCCCAC	A CAMCA CMC	CCTTGCCAAG	TACATCAACG
1			CTCCTCCCCA	A A COMO A COM	AACATGACTC	CCGAGTTCCT	CGACTACGTG
1		i	CIGCIGGGCA	AAGGTGAGGT	TGGGAAAGTT	GCGGAAGCTT	CAAAACTCAG
1		1	CGIGGAGTIA	ATCCAGCAGA	TCAGGGACGA	CTTCGAGTAC	TGGTATCCCG
			TIGACCTAAG	AAGCAGTGGC	AAGGACTTGG	TTGCAAACCA	CCTGCTCTTC
			TACCTCTTCC	ACCACGTCGC	CATTTTCCCG	CCAGATAAGT	GGCCGAGGGC
1			AATTGCCGTA	AACGGATACG	TCAGCCTTGA	GGGCAAGAAG	ATGAGCAAGA
1			GCAAAGGGCC	CTTGCTAACG	ATGAAGAGGG	CGGTGCAGCA	GTATGGTGCG
			GATGTGACGA	GGCTCTACAT	CCTCCACGCT	GCAGAGTACG	ACAGCGATGC
			GGACTGGAAG	AGCAGAGAGG	TTGAAGGGCT	TGCAAACCAC	CTCAGGAGGT
1			TCTACAACCT	CGTGAAGGAG	AACTACCTGA	AAGAGGTGGG	AGAGCTAACA
1			ACCCTCGACC (GCTGGCTTGT	GAGCAGGATG	CAGAGGGCAA	TAAAGGAAGT
		ĺ	GAGGGAGGCT	ATGGACAACC	TGCAGACGAG	GAGGGCCGTG	AATGCCGCCT
		İ	TCTTCGAGCT (CATGAACGAC	GTGAGATGGT	ATCTGAGGAG	AGGAGGTGAG
1			AACCTCGCTA	TAATACTGGA	CGACTGGATC	AAGCTCCTCG	CCCCCTTTCC
1			TCCGCACATT	TGCGAGGAGC	TGTGGCACTT	CAACCATCAC	ACCURACCION
		j	GCCTCGAAAG	TACCCAGA	TACGACGAAA	CCACCCMMCA	CCACCACCCC
-			GAGAGAATTG		CCGAAACCTT	CCAGGGIIGA	CGAGGAGGCG
			CAAGAAGTTT	TOURTHCCI	CGAAGGAGGT	GITGAGGACA	TTCAGGAAAT
1			ACTGGAACCT O	DAYCCCACCA	A A COMOCHEC	TTACATTGCT	CCCGCCGAAG
			ACTGGAAGGT T	ACCA COMMAN	AAGGTCGTTG	CTGAAAGCGG	GGATGTTGGG
			GAGGCGATGA	AGCAGCTTAT	GUAGGACGAG	GAGCTTAGGA	AGCTCGGCAA
ł			AGAAGTGTCA	AATTTCGTCA	AGAAGATTTT	CAAAGACAGA	AAGAAGCTGA
1		İ	TGCTAGTTAA (GAGTGGGAA	GTTCTGCAGC	AGAACCTGAA .	ATTTATTGAG
			AATGAGACCG G	SACTGAAGGT	TATTCTTGAT A	ACTCAGAGAG	TTCCTGAGGA
F	EO ID 14		GAAGAGGAGG C	CAGGCAGTTC	CGGGCAAGCC (CGCGATTTAT (GTTGCTTAA
3	EQ ID: 14	Methanob	GTGGATATTG A	AAGAAAATG	GCGTGATAGA '	TGGAGAGATG (CTGGCATATT
	İ	acterium	TCAGGCTGAC C	CTGATGACA	GAGAAAAGAT A	ATTCCTCACA (GTCGCTTACC
		thermoau	CCTACCCCAG I	GGTGCGATG	CACATAGGAC A	ACGGGAGGAC (CTACACTGTC
		totrophi	CCTGATGTCT A	TGCACGGTT	CAAGAGGATG (CAGGGCTACA	ACGTCCTGTT
		cum	TCCCATGGCC T	GGCATGTCA	CAGGGGCCCC	TGTCATAGGG	ATAGCGCGGA
		leucyl	GGATTCAGAG G	AAGGATCCC	TGGACCCTCA A	AAATCTACAG	GAGGTCCAC
}		1	AGGGTCCCCG A	GGATGAGCT	TGAACGTTTC Z	AGTGACCCTC 1	AGTACATACT
		tRNA-	TGAATACTTC A	GCAGGGAAT	ACCGGTCTGT T	PATGGAGGAT 2	ALCCETAGE TO THE PROPERTY OF T
	İ	syntheta	CCATCGACTG G	AGGCGTGAA	מאטממממטייי ל	CCDACCONC C	TOGGCIACI
	j	se	TTCATACAGT G	CCACATAAC) JAJJJIIADU.	TACAGCAGG
	-	(MtLRS)	GGGCGCCCCAM C	CUCUUN YOU COCCUINTANG	CULCCCUCA *	TACCTIGGCC 1	CGTAAGGAA
		,	GGGCGCCCAT C	CIGIIAAGT .	ACTGCCCTGA A	AIGIGAAAAC (CTGTGGGTG
_		<u></u>	ACCATGACCT C	CTTGAGGGT (GAGGGGGTTG C	CATAAACCA G	CTCACACTC

CCCARATTCA ARCTTCGAGA CTCATACCTG GTCGACCCA CTCTCAGGCC CGAGACANT TATGGGGCA CCAACCTCTG GCTGAACCCT GATGAGATA ATCTCAGGGT TGAAACAGT GGTGAGGACT GTGGATATAT TTTCACACA GAAACTGGAC CTCAAGGTT CCGGTGAGCAC CAACCCCGGG GACCTGATAG GAGACTGCT GAGAGATCT GGGATATAT TTTCACACA GAAACTGGAC CTCAAGGTT CCGGTGAGCAC AAGACACCC CATACTCCCG GGTCTCTTCTG TTGACCTGA ATATGCCACA GGACACCCC AGGACAGACC TGAAAGTGC CCTGCAGAT TCATAGCCCT TAGGACCTC AGGACAGACC TGAAACTGC CCTGCAGACT TCATAGCCCT TAGGACCTC AGGACAGACC TGAAACTGC CTTCACCAGT GAGAGGTCA GGTGAGTTCC CGCCTGAGG CCGCTGAGT TCATAGCACT GGAGAGCTC AGGAGACCCC CGCCTTGAGG ATGCCACCG GGAGGTATA AGAACTGAC GGAGACCCC CGCCTTGAGG ATGCCACCG GGAGGTATA AGAACTGAC GGAGACCCC CCCCTGAGG ATGCCACCG GGAGGTATA AGAACTGAC GTCTCTCAGG CCCGTCAGGG GAGCTATATA AGAATCGAG GTCTCTCAGG CCCGTCAGGG ATCCTCATGC GGAGGTATA AGAACTGAGA GTCTCTCAGG CCCGTCAGGG ACCACTCC CTGCTATATA AGAACTGAG GTGGCAGGT CTTGTGAGG CTCTGCACAG ACCACTCTC ATGCGAGAG TTCAATGACTG CCTGGAAGGA CCTGCCCCA AGCTCCCTG ATGCAAGAG TTCAATGACTG GCCATGTATCA AGAACGAGA ACCACTGCAC GCATATTACC GAGAGGTC GCCACATTCC ATGCATGAC TCAATGACTG GCCATTATCA AGGAGGATAG CCCTGCAGGA GGAGATCTC ACCACTCCCA ACCCCCTCA AGGACTCAA CAATCTACAT TGGATGAGA GCCATTTCA ACGACCCCCC ACCACCCCTCA AGCACCCAA GGCATATTAC ACCATTCCAC ACCACCCCCTC AGGACTCAA CAATCTACAT ACCTCACAAGA GCCCTCTTTGATG GCAATCACCT GGAGATCTA ACCATTCCAC ACCACCCCTCA GCGACTTCAA CAATCTACAT ACCTCACAAGA GCCCTCTATAG GCAATCACCT ACCTCACAAGA CCCCCTCAAGACCCCCC ACCACCCCTCA ACCACCCCTCA ACCACCCCTCA ACCACCCCTCA ACCACCCCTCA ACCACCCCCTCA ACCACCCCTCA ACCACCCCTCA ACCACCCCTCA ACCACCCCTCA ACCACCCCCC CAACCACCCCCC ACCACCCCCCTCA ACCACCCCCCC ACCACCCCCC CAACCACCCCCC CAACCACC							
ARGERGAGEMENT TRICARCACE GARACTIGAC COTAMOGRATIC CORGANIZATION CONTROLL AGARACTOCA GARACTICAC CONTROLL AGARACTOCA GARACTICAC GARACTICAC GARACTICAC TRACAGACCA CATACTCCCA GATACTCCCA GATACTCCCA ATTACTCCCA GAGACACCAC TOTACAGACCAC TACCAGACCAC COCCTOTAGA ATTACTCCCAC COCCTOTAGA ATTACTCCAC GAGAGACCCC COCCTOTAGA ATTACTCCACACACACCCC CACCTATAC ARACTACCAC ATTACTCCACACACACCCCC CACCTATACACACAC			CTCAAATTCA	AACTTGGAGA	CTCATACCTG	GTCGCAGCCA	CCTTCAGGCC
ARGERGAGEMENT TRICARCACE GARACTIGAC COTAMOGRATIC CORGANIZATION CONTROLL AGARACTOCA GARACTICAC CONTROLL AGARACTOCA GARACTICAC GARACTICAC GARACTICAC TRACAGACCA CATACTCCCA GATACTCCCA GATACTCCCA ATTACTCCCA GAGACACCAC TOTACAGACCAC TACCAGACCAC COCCTOTAGA ATTACTCCCAC COCCTOTAGA ATTACTCCAC GAGAGACCCC COCCTOTAGA ATTACTCCACACACACCCC CACCTATAC ARACTACCAC ATTACTCCACACACACCCCC CACCTATACACACAC	į		CGAGACAATC	TATGGGGCCA	CCAACCTCTG	GCTGAACCCT	GATGAGGATT
GRGGATAATC TTTCACACCA GAAACTGGAC CTCAGAGGTTT CCGGTGAGGAC AGAACACCC GACCCTATAG GAGATGCCCT GGAGAATCCT GTGACGGGCC AGGACACCCA GACACCCCCCCCCTCAGAGCC ATTATAGCCACA GGTGTTGTGT TCTCTGTCCC TGCACATGCC CTGCACAGCT TGATAGCACCA GGTGTTGTGT TCTCTGTCCC TGCACATGCC CTGCAGAGCT TGATAGCACCA GGTGGTTGCCCCCCCCCCTGAGG GGTTATAGAG GGTGGTTGCCCCCCCCCCTGAGG GGTTATAGAG GGTGGTTGCCCCCCCCCCCCCTTGAGG GGTTATAGAG AACTCAAGCCACCCCCCCCTTGAGG GGTTATAGAG AACTCAAGCCACCCCCCCCCTTGAGG GGCCAACCC CTCCTAGAGCCCCCCCCCC		·	ATGTGAGGGT	TGAAACAGGT	GGTGAGGAGT	GGATAATAAG	CAGGGCTGCC
CAACCCCGGG GACCTGATAG GGARTTCCT GTGACCGGCCC AGGACCACC CATACTCCCG GCTTCCTTGC TGACCCTGA ATTAGCCACA GGTGATTCTC TCAGACAGACACC TGAGACTTC TGAAAGTACCCT TGAAAGTACCCT TGAGGACTCC GGAGCAGACC ATTAGCACTA ATTAGCTTC TGATATTCAG CCCGGAATGC CCTGCAGACT TCATAGCCCTA GGAGGACCC CGCCTTGAGG ATGCCACCG GGACCTATAC AGAATCGAC GGTGACGTCC CGCCTTGAGG ATGCCACCGG GGACCTATAC AGAATCGAC ATGCGAGGG TGTTATGAGC AGCCACATCC CTGTCTATGG TCAGAAACCA ATGCGAGGG TGTTATGAGC AGCCACATCC CTGTCTATGG TCAGAAACCAC TGCCAGACGG CCGTGTGAGG ATGCCACTGA GACCACTACAC GTGCACAGGG CTGTTGAGG ATGCCACTGA GACCAGTGGT TCAGAGGACGC TTCTATGAGG CCGTGAGGG ATCCTCCCCAC AGGTCCTTATA TCCCGGTCAG GTGACAGGG CTGTGAGGG ACCTGCCCAC AGGTCCTCA AGGACCACT TCAATGACC GCACTTTCTA AGGAGGATAG CCTTGAAACCAC TGGACATATAC ACCATCGCAC ACCGCCTCA GGACACTCAT TGGACAGTACA ACCATCGCAC ACCGCCTCA GGACACTCAT TGGACAGTACA GGACATCTTC AGCAGAGATAG CCTTGGAAC ACCGCTCAC GGCATATTAC ACCATCGCAC ACCGCCTCA GGACACTGAGA ACCACTCACAC AGGACACTCA ACCATCGCAC ACCGCCTCAG GGAATCACCC GAGGACTCAA GCCACTCTTCTAGAC ACCACTCACACACACCCTC ACGACCCTC ACTACACCAT ATTCCCTGAG TCAGGAGGAT TCCCCCTAT ACCCCCTTT ACTCGAGGACACACTC ACTCACACCAT ATTCCCTGAG TCAGGAGACA CCCGGGGGCA ACCAGGGCT TCTTCTAGGA GACCTATTACC TCCCCCAC GGACACCTC ACTCCACCCTG ACTGAAGACC TCAGACACCC CCGGGGGCACTTACCCCCTAC ACTGCACACCC CATACTCCTCA GGCAACCCC TCAGGACACC TCCCCACCTCCA ACGCCACCTC AGGCAACCTT CTCTCACCA GAGCCATCTCCCA ACGCCACCTC AGGCAACCTT CTCTCACCA GAGCCATCTCCCA ACGCCACCTC AGGCAACCT TCCCCCACCCC CAGGGAGAT TCCCCCTACTCCA ACCCCAGAGACA CCGTCCACCTC ACTGAAGCC TCTACTCCCA ACCCCAGAGCACC TCAGACACCC CCGGGGGCACCT TCACCCCCACCCC ACCAGAGGACA CCGGCCACCTCACCCC ACCAGGAGACA CCGCCACCCC ACCAGAGCACCA TCCCACCCCC ACCAGAGGACACCA TCCCACCCCC ACCAGAGGCC CTCATCTCCCTACCCCACCCC		ŀ	GTGGATAATC	TTTCACACCA	GAAACTGGAC	CTCAAGGTTT	CCGGTGACGT
AGGACACCC CATACTCCC GETTCTTGT TGTATCCCTAC GETTCTTCTTCTCT TCACATCTCC CTCACACACT TCATAGCCT TGAGGACTC TGAGGACTC AGGACAGACC ATGACTTCT TGAAAGGTAC GGTGAGTTCC CGCCGCCGA GCTTATAGAC ATGTGGTCC CGCCGCCGA GCTTATAGAC ATGTGGACCC CGCCTGAGAT TCATAGCAG GGTGAGTTCC CGCCGCCGA GCTTATAGAC ATGTGGAGGG TGTTATAGAC ACCACTCC CTGCTATAG ATGCGAGGG TGTTATAGAC AGCACACTC CTGCTATAG TAGAACCA ATGCGAGGG TGTTATAGAC AGCACTCC CTGCTATAG TAGAACCA ATGCGAGGG TGTTATAGAC AGCACACTC CTGCTATAG TAGAAACCA ATGCGAGGG TGTTATAGAC AGCACACTC CTGCTATAG TAGAAGCAC GTCTCTGAGAG ACCAGGGTC TAGACTAGAC AGAACGACCA GTCATATACCC GAGGAGGAT TCCACAGACG ACCAGTGGTT TCACATAGACG CCTGGAAGGA CCTCGCCCAC GGCATATTAC CGCATGAGA CCTCGCCCAC GGCATATTAC ACCATGCCAC ACGTCCCTCA AGGACCACAC GGCATATTAC ACCATGCCAC ACGCCCCTCA CAGACCACAC GGCATATTAC ACCATGCCAC ACGCCCCTCA CAGACCACAC AGGACTCTCA GGCAGATATTC CAACCACCCC ACGCACCACA AGGCCTGCAC TGCACCATCA GCCATCATAG GCAATCACCT GAACTCACAC ACCTCACACCA GTCCTTTCTA GCCATATTCC TAGACTACAC ACCTCACACCA GTCCTCTATAG GCCATACACCT GAACTCACAC ACCTCACACCA GTCCTTCTAG GCCATACACC GAACTCACAC ACCTCACACCA GTCCTTCTAG GCCATACACC GAACCACTCAC ACTCCACCAC ACCTCACACCA ACCCCCCTGA GCAACACACAC ACCTCACACCA ACCCCCACACACCACAC			CAACCCCGGG	CACCTGATAG	GGATGTGCGT	GGAGAATCCT	GTGACGGGCC
GGNTTGTTCT TCTCNTTCC TGCACATGCC CTGCAGACT TCATAGCCCT TGAGGACCC AGGACAGCC AGGACACCT TGAAAGCTAC GTTTGAGG GGNGAGTTCC CGCGGCGCGA GGTTATAGA CAATTTGGT TCAGAGCACG GGNGAGCCC CGCCTGAGG ATGCCACCG GACCTATCA CAAGACCAA GGAGGACCC CGCCTTGAGG ATGCCACCG GACCTATCA CAAGACCAA GGAGGACCC CGGCTTGAGG ATGCCACCG GACCTATCA CAAGACCAA GGTGCAGGT CTGCAGAGGA CCCGTGAGGA CCTGTCATAG AGACCAA GGTGCACGG CCGTGAGGA CCCTCCACCAC CGGCTTATAA TGCCGTGAGG TCTCATGAGG CCGTGAGAGA CCTCGCCCAC AGGTCCTTATA TGCCGTGAGG CCTTGAATGAC GCTGTGAGG GCACTGAGGA ACCTGTTATA TGCCGTGAGG TCTAATGACC GGCATTTCA AGGAGGATAC ACCTGTGACACACACACACACACACACACACACACACACA		i	ACCAACACCC	CATACTCCCC	CCTTCCTTCG	TTGACCCTGA	ATATGCCACA
TGAGGACCTC AGACAGACC ATGAACTCT TGAAAGGTAC GGTCTTGAGG ATGCGTACC TGATATTAGA CCCCTGAGAAT TCATTAGCAGT GGATGCTAC GGAGGACCCC CGCCTTGAGG ATGCCACCGG AATGCGAGGGG TGTTATAGA AAATTTGGTG TCAGAAACCA ATGCGAGGGG TGTTATAGA AAATTTGGTG TCAGAAACCA ATGCGAGGGG TGTTATAGA ATGCACCGG GGACCTATAC AGATCAGGC ATGCCAGGGG TGTTATAGA ATGCACCGG GACCATTC TCAGATGAG CCCGTGAGGT CATCGCTGAT GAACTGAGAG ACCAGGGCCT TCAGATAGAC GCGTGAGAGA CCTCGCCCAC GGTAGATACCC GAGGAGGTC CGGCCAACTT TGAATACTAC ATCGATGAG GATAATACCC GAGGAGGTC GGGCCAACTT TGAATACTAC ATCGATGAG GGCATATTAC ACCATGCAC ACCGCCTCA GAGGAGCTAC ATGCATGAGA GGCATATTAC ACCATGCAC ACCGCCTCA GAGGAGCTAC ATGCATGAAG GGCATATTAC ACCATGCAC ACCGCCTCA GAGGAGCTAC ATGCATGAGA GGCATATTAC ACCATGCAC ACCGCCTCAC GGAGCACACA TGGACGAGA GGTGACTACT GGACCCCCT ACGAGCTACA AACCTGACA GGCATATTAC ACCATGCAC ACCGCCTCA GAGAGCTACA AACCTGACAC ACTCAGCAGA GACCACTATGA GCATACACCT GACATCACAC ACTCAGCACA ATTCCTCTAGA GCATACACCT GACATCCACA ACCGCCTCAC GGATAGGCC TCTCTCAGCA ACCCCTCAGA GACCACTACACACACACACACACACACACACACACACACA	1.		AGGAACACCC	CATACICCCG	MCCA CAMCCC	CCTCCACACT	ጥሮልጥልሮሮሮሮጥ
ARTGGTTGC TGCATATTGAG CCGTGAATT GCATAGCAG GGATGGATGGATGGATGGATGGATGGATGGATGGATG	1		GGTGTTGTGT	TCTCTGTCCC	TGCACATGCC	TORDADOLOG	CCHINGCCCI
GGTGAGTTCC CGGCGGCGCA GGTTATAGAG AAATTTGGTG TCAGAAACCA GGAAGACCCC CGCCTTGAGA TCACCACCGG GAGCTATAC AACATCGAGC ATGCGAGGGG TGTTATGAGC AGCCACACCC CTGTCTATGG TGGTATGAGA GTCTCTGAGG CCCGTGAGGT CATCGCTGAT GAACTGAAGG ACCAGGCCTT TGCAGATGAGA ATGTATGAAT TGCCTGAAGG ACCAGGCCTT TGCAGATGAG CCTGGAAGGA CTCCGCCAC AGGTGCCTG ATGCCATGAG GTGGCAGGTG CGTGTGAGG GTCATCGCCAC AGGTGCCTG ATGCATGAGA GATAATACCC GAGGAGGTCC GGCCAACTT TGAATACTAC ATGCATGAG GGATATTAC ACCATCGCAC AGGCCCTCA GAGACTCAC AAACTCACAC TCAATGACTG GGCATATTCA AGGAGGATTA CACATCGCAC AGGCACTAC GGCATATTAC ACCATCGCAC ACCGCCTCAG GGAGATCAC CACAGCTCAC ACAGCTCACA TGGAGCAGAA GTTCTTCATGA GCCATATTCC TACAGTAGTC GGAGATCCAC GGCATATTCC TACAGTAGTC CCGGGAGACCT GGAGATACCAC GGCCCTCTATAG GCAATTCCACT GGAGATCCAC ATTCCCTCTGAG TCACGTATTCC TACAGTAGTC TCATGAGACCTT CTCTCCACACA ATTCCCTCTGAG TCACGTACTCCAC ACTCACCAC ACTCACCCCAT ATTCCCTCTGAG TCACGAGAGCAC CTGTGCACAC ACTCACCCCTA ATTCCTCTGAGG CACAAGAGTT CACTCCACCA AGGCACACTT CATACACTCA TCCGGAGCCC CAGAGAGCA CGGTGCAGAC CTGTGCCCC CATACACTCCT AGGGAGCCC CAGAGAGCA CGGTGCAGAC CTGTGCCCC AAGAGGGTT CTGCGACCC CAGAGAGCA CGGTCCACA AGGCAATCGA ACGAGAGGA TACGTGAACC ACCAAGGACC CAGGCACTTCA CTGAGGGGATC AACCACAGGA TACGTGAACC ACCAAGGACC AGGCATTCAC CTGAGAGCAC AAAGGCAGTT CAGGAGCAC TCTATTCTCAT ACCTTAAGCGT TCAGCAGACC ACAAGGGCC TCTATCTCCT AACCACAGGA TACGTGAACC ACAAGGGCC TCTATCACCA AGGCAACACA AAAGCACTT CAGGAGCAC TCTATCTCCT TAAAAAGGAT TCCAGCACAG AAAGCACTCAT GCCCCACACC CAGAGGACCA GCGAGAGCACA AAAGCACTCAT GCCCCCC GATAAGCC TCAAAGGGC TCTACACCACA AAAGCACTCAT GCCTCACCC CGACAGGCCC CACAGAGCAC CACAGGACACA CACACCACAG GCCCCC CAGAAGCAC CACAGGACCAC CACACACA	İ		TGAGGACCTC	AGGACAGACC	ATGAACTCCT	TGAAAGGTAC	GGTCTTGAGG
GGAGGACCCC CGCCTGAGG ATGCCACCGG GGAGCTATAC AGAGTGAGAGG ATGCAGGGG TGTTATGAC AGCACATCC CTGTTATAG GGGATGAGAG GTCTCTGAGG CCCGTGAGGT CATCGCTGAT GAACTGAGG ACCAGGGCCT TGCAGATGAG ATGTATGAAT TGCCTGAGGG ACCAGTGCTT CATGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	- i		ATGTGGTTGC	TGATATTGAG	CCCGTGAATG	TCATAGCAGT	GGATGGCTAC
ANGCGAGGGG TCTTATAGAGC AGCCACATCC CTGTCTATGG TGGTATGAGGCGCT TCCAGATGAG CCCTGAGGGT CATCGCTGAT GAACTGAGGG ACCAGGGCCT TCGAGGAGGG CGTTGTGAGG GTCATGGAGG ACCAGTGCTT CAGAGTGAG TCTGATGAGG CCTGGAAGGA CCTCGCCCA AGGTGCCTCG ATGGATGAGG GATAATACCC GAGGAGGAC CCTCGACCCA AGGTGCCTCG ATGGATGAGA GATAATACCC GAGGAGGAC CCTCGCCCA AGGTGCCTCA TCGAGTGAGA GCCTGATTCA AGGAGGATAG ACCACTGCAC AGGCTCACA GGCATATTAC ACCATCGCAC AGGACCCTC ACAGACTCAA AGGCTGCCC AGGACATTTCA ACCATCGCAC ACGCCTCAG GGAGATCACA CGGCGATATTCA GGCATATTCC TAGATACTTC AGGAGCCTT GAGGATCTCA GGGAGGAATT CCGGTAATTCC TAGATGATATC ACTGGACGCT ACTCAGCCAT ATTCCTTGAG GCAATTCCAT TAGATGATACA GGTATGCGCC TTCTTGAGG CAACAGAGT TCCCCCGGGGGG TCTGTGCTCTAGACCCAT ATTCCTCTGAT TCAGAGGTGCC CAGACCCCT AGGCCACATTCC TAGATCCTCA AGGCACCTTT GGTATCCCTG GGGAGGAATT CCGGTAATTCC TAGATGATTCA ATATTCCACA ACCTCACCCAT ATTCCTCTGAG CAACAGAGGT CAGACCCAT TCTCTCAGAG CACCACAGGGGGACCCATCTCTCAGACCACACAGAGCA CGGTCCACACACGCCCATCCCA AGGCAACCATCACCACACACACACACACACACACACACAC			GGTGAGTTCC	CGGCGGCCGA	GGTTATAGAG	AAATTTGGTG	TCAGAAACCA
ANGCGAGGGG TCTTATAGAGC AGCCACATCC CTGTCTATGG TGGTATGAGGCGCT TCCAGATGAG CCCTGAGGGT CATCGCTGAT GAACTGAGGG ACCAGGGCCT TCGAGGAGGG CGTTGTGAGG GTCATGGAGG ACCAGTGCTT CAGAGTGAG TCTGATGAGG CCTGGAAGGA CCTCGCCCA AGGTGCCTCG ATGGATGAGG GATAATACCC GAGGAGGAC CCTCGACCCA AGGTGCCTCG ATGGATGAGA GATAATACCC GAGGAGGAC CCTCGCCCA AGGTGCCTCA TCGAGTGAGA GCCTGATTCA AGGAGGATAG ACCACTGCAC AGGCTCACA GGCATATTAC ACCATCGCAC AGGACCCTC ACAGACTCAA AGGCTGCCC AGGACATTTCA ACCATCGCAC ACGCCTCAG GGAGATCACA CGGCGATATTCA GGCATATTCC TAGATACTTC AGGAGCCTT GAGGATCTCA GGGAGGAATT CCGGTAATTCC TAGATGATATC ACTGGACGCT ACTCAGCCAT ATTCCTTGAG GCAATTCCAT TAGATGATACA GGTATGCGCC TTCTTGAGG CAACAGAGT TCCCCCGGGGGG TCTGTGCTCTAGACCCAT ATTCCTCTGAT TCAGAGGTGCC CAGACCCCT AGGCCACATTCC TAGATCCTCA AGGCACCTTT GGTATCCCTG GGGAGGAATT CCGGTAATTCC TAGATGATTCA ATATTCCACA ACCTCACCCAT ATTCCTCTGAG CAACAGAGGT CAGACCCAT TCTCTCAGAG CACCACAGGGGGACCCATCTCTCAGACCACACAGAGCA CGGTCCACACACGCCCATCCCA AGGCAACCATCACCACACACACACACACACACACACACAC	ŧ		GGAGGACCCC	CGCCTTGAGG	ATGCCACCGG	GGAGCTATAC	AAGATCGAGC
GTCTCTGAGG CCGTGAGGT CATCGTGAT GAACTGAGGG CACGGGCT TCTGATGAG ATGTATGAAT TCGCTGAGG ACCGTTATA TGCCGCTGGG GTGGCAGGTG CCTTGTGAGG CTCGCCCAC ACGGTGCT CATGAGTAC CCTGGAAGGA CCTCGCCCAC AGGTGCCTCG ATGGCATGAA GATAATACC GAGGAGTC GGCCAACTT TAGATGCAC ACGCTGCCC TCAATGAGA GGTGATCAT AGGAGGCAACTT TGGACGATGA GGCAGTTTC TGAATACTAC ACGACTGCC GGCATATTAC ACCATCGCAC ACCGCCTCAG GGCAGTGAC AGACTGCTC GGCATATTAC ACCATCGCAC ACCGCCTCAG GGCAGTGAC AGACTCCTC GAGGATCAC AGGAGAATT CGCATCACCAC ACAGACTCAC ACAACTCACAC ACTCAGCCAT ATTCCCTGAG GACCACTCT ATGATGATTC AGGAACCTTT GGTATGAGCCAT ATTCCCTGAG GACCACTCT ATGATGATTC AGGAACCTTT GGTATGAGCCAT ATTCCCTGAG GACCACTCA TAGATGATTC ACGAACCTTT GGTATGAGCCAT ATTCCCTGAG GACCACTCCA ACATCCACA ACTCACCCAC ACACCTCCA ACCGCCACACCTCACACACACACACACACACACACACACA	1		ATGCGAGGGG	TGTTATGAGC	AGCCACATCC	CTGTCTATGG	TGGTATGAAG
GTGCAGATGAG ATGTATGAGT TCGCTGAGG ACCGGTTATA TGCCGTGCGCAC GTGGCAGGTG CGTGTGAGG CCTGCCCAC AGGTGCTC CATGAAGTAC TCTATGAGC CCTGGAAGGA CCTCGCCAC AGGTGCTC ATGGATGAC TCAATGACTG GCATGTTCA AGGAGGATAG AGGACTGGCAC TGGATGAGA GTGCATTTCA AGGAGGATAG CCCTGGAACAC AGAGCTCAC AGAGCTCAC CGCATATTCA GGCATTTCA AGGACTCAC CACACTCACA CAACACTCAC CACACTCACACACA	ŧ		CTCTCTCAGG	ССССТСАССТ	CATCGCTGAT	GAACTGAAGG	ACCAGGGCCT
TCTGATGACG TCTGATGACG TCTGATGACG GCTGGAAGGA GCTCGCCCAC GAGGACGTCC GAGGAGGACCCTC TCAATGACTG GCATATTAC GAGGAGGACCCTC TCAATGACG TCGAATGACG GGCATGATCA GGCATCATC GGCATCCC GGCATCATC GGCATCCC GGCACACTT TGAATGACT GGCATCCC GGCACACT TGAACCCTC GGCATCATC GGCATCCC GGCACACT TGAATCACCC CTCTGCAAG ACTCACCCAT ACTCCTCAG ACTCACCCAT GGCATCCC CATACTCCTC AGGGATGCC CATACTCCTC AGGGATGCC CATACTCCTC AGGGATGCC CATACTCCTC AGGGATGCC CATACTCCTC AGGGATGCC CATACTCCTC AGGGATGCC CATACTCCTC AGGGATGCC CATACTCCTC AGGGATGCC CATACTCCTC AGGGATGCC CATACTCCTC AGGGATGCC CATACTCCTC AGGGATGCC CATACTCCTC AGGGATGCC CAGGAGGACC CTCTCAGCCA AGCGCATCCT CTCCAGCTA ACTCAGCACA ACCTGAAACC TCACTGCTGA ACCTGAAACC CTCCAGCTAA AACCAGAGGA AACCAGAGGCA CTCATCTCCC AAACGCTC CACAGAGGCA CTCATCTCCC CACAGAGGCA AAACGCATT CACGGAGCCC CACAAGGCC CACAACGCTC CACACACACAC AAACGCATC CACACCCTC GACATCTCC CACACCCTC GACATCTCC CACACCCTC GACACCTCC CACACCCTC CACACCCTC CACACCCTC CACACCCTC CACACCCTC CACACCCTC CACACCCTC CACACCCTC CACACCCTC CACACCCTC CACACCCTC CACACCCTC CACACCCCTC ACACCCCC CACACCCCC CACACCCCC CACACCCCC CACACCCCC CACACCCCC CACACCCCC CACACCCCC CACACCCCC CACACCCCC CACACCCCC CACACCCCC CACACCCCC CACACCCCC CACACCCCC CACACCCC CACACCCC CACACCCC CACACCCC	į		TO CACATOA	ATCTATCAAT	TCCCTCACCC	ACCTGTTATA	TGCCGCTGCG
TCTGATGAGG CCTGGAAGGA CCTCGCCCAC AGGTCCTCC ATGGCATGAA GATATACCC GAGGAGGTCC GGCCAACTT TAAAACTCA ATGGATGGC TCAATGACTG GGCATGTTCA AGGAGGATAA CACTATCACT GGGATGAAG GGTGGATCAT CGAACCCCTC ACAGACTCAA CAATCTACAT GGCATATTAC ACCATCGCAC ACCGCCTCAG GGCAACTCAT ACAGACTCAA TGGACGATGA GTTCTTTGAT GCCATATTCC TAGATGGAT GCCGGGGAAG TGGACGATCAA GGACCTCATAG GCAATCACTG TACCTCATA ACTCAGCCAT ATTCCCTGAG TCAGGGAGGAT ATATTCCACT GGTATGGCC TTCTTCAGGG CAAACACGT GACATTCCAT ATATTCCACC CATACTCCTC AGGGATCCCA TCAGGGAAGCA CCGTGAGGCAC ACTCACCCAT ACTCCCTCAGG CAACAGATG CCATCACTCA AGGGAACGCA TCTTCCTCAT GTCCTCAGGG CAACAGATG CCATCACCTCA AGGGAACGCA AGGAGGCTC TCAGGGACCG CAGGAGGATT GAATGGTCAG AGGAGGGT ACGGGACCGC CAGGAGGATT GAATGGTCA CGGGAGGAT ACCAGAGGA TACCTCAAGAC AAGCAGGA TACCTCAAGAC AAAGCAGGA TCCTCAAGAC TTCATTGCAA GGGAATTCAG AAACAGAGGAT TCAGGAGACCA CTCATCTCCA GAGGAATTCA AACCAGAGGA TACTGAAGAC TTCATTGCAA GGGAATTCAG GGAATTCAG AAACCAGAGGA TACTGAAGAC TTCATTGCAA GGGAATTCAG GGAATTCAG AAACCAGAGGA TACTGAGAC CACAAGGGCC CTCAAAGACC TCACAAGGAC TCTATTCCT TAAAAAGAT TCCAGACAAG AACCACAGGA TACTGAGAC CACAAGGGCC CTGAATCAT TCCAGACAAG AACCAGAGGA ACCTTAACCCTT CACAAGGAC TTTCATCACCTA ACCTTAACC TCTCACCCC GAATAGGCT ATGGGTAACA CTTCACCCAC AAAGCTCCAT GCCACCCCT GATAAGGCT ATGGGATAAT GTTGCACCAT CCACACGTT GCACCCCT GAATAGCT TAAAAGACT TCAACACATA AAGCTCATG GCCTGACTT CAACAAGGAC TTGTACCACTACACACACACACACACACACACACACACAC			CTCCAGAIGAG	AIGIAIGAAI	CMCAMCCACC	እርር ነር ተጠመተ እርር እርጥር ርጥጥ	CATCAACTAC
GATAATACCC GAGGAGGTCC GGGCCAACTT TGAATACTAC ATGCATTGGC TCAATGACTA GCATGTCCA AGGAGGATAG AAGCCTCCCC TGGGATATAC GCCATGTCCA CACACCCCC ACAGACTCAA CAATCTACAT GGCATATTAC ACCATCGCAC ACCGCCTCAG GAGACTCAA CAATCTACAT GAGGATCACA GTTCTTTGAT GCCATATCGC TACAGCATTC AGGAACCTT GAGGATCACA GTTCTTTGAT CCCTACTG TACCCCTTG ACGACCTCAT ACTCACCCAT ATTCCCTCAG TCAGGTGTGC CCCGGGGGATA ACTCACCCAT ATTCCCTCAG TCAGGTGGC CCCGGGGGGC TTCTTTGAGGG GGTATGGCC TCCTCAGCA GAGCATTCCAT AGGAACCTCT AGGAGGATCA TTCCTCAG TCAGGAGGAT TCATCCTCA AGGCCATCGT AGTAGGCC TCCTCAGCA GAGCCATGC CCCGGGGGGC TTGGGTCTTT CTCTCCTAT GTCCTCAGCA GAGCCATGC AGGCATTCAA AGGCAACCT AGTGAGGGCA TCCTCAGCA GAGCCATGC AGGCATTCAC AGGCAACCA AGTGAGGGTC TCAGGAACCA TCCATGAGC AGCCATTGC AGGGAATCGG AGAGAGGGTC TCAGGAACCA TCTATTGCAAC AGGCAATCCT AACCTAAGCG TCCTGAAACC TCTATTCCAA AGGAATCAC AACCAGAGGA TACGTGAAGC TCTATTCCAA AGGATTCA GGGAATCGG AAACCAGTGA TACGTGAAGC TCTATTCCAA AGGATCACT AGCTTAAGCC TCCAGCAAGA AACCAGAGGA TACGTGAAGC ACCAAAGGGC CTTGAATCAT TACAGCACAGA AACCAGAGGA TACGTGAAGC ACCAAAGGGC CTTGAATCAT ACCTTAAGCC GCAAACGTTC TGCACCACAA AGGACATTC AGGATCACCT AGACTTAAGC TCCAGCACAG AAGGCTATG GAAAGGACATC AGGCTCAT TACACCTAA AGGACATTC AGGCTCAT TACACCTAA AGGACATCAC GAAAGACATC AGAGCATCAC CACAGAGGAC AGAGCATCAC CACAGAGGAC AGAGCATAC AGGCTCACT CACACACAC AAGGCACAC AGAGCATCA CACACACACACACACACACACACACACACACACA			GTGGCAGGTG	CGTTGTGAGG	GTCATGGAGG	ACCAGIGGII	AUCCCAUCAA
TCAATGACTG GGCATGTTCA AGGAGATCA CAATCTACA CAGCTCACA GGCATATTAC ACCATCACA ACCATCACA CAGCTCACA GGCATATTAC ACCATCACA ACCATCACA CAGCTCACA GGCATATTAC ACCATCACA ACCATCACA CAGCTCACA GGCATATTAC ACCATCACA CACCATTACACT GCCATATTAC ACCATCACTACA GGCATATTAC ACCATCACTACACTA			TCTGATGACG	CCTGGAAGGA	CCTCGCCCAC	AGGIGCCICG	ATGGCATGAA
TEGGATGAGA GGTGGATCAT GCAACCCCTC ACGACTCAAC GCGAGGAGA GGCATATTAC TGGACGATGA ACCATCGCAC ACCGCCTCAG GGAGAGTGAA CCATCAGCAC TGGACGATGA GTTCTTTGAT GCCATATTCC TAGATGATTC CTCTGCAAAG GACCTCATAG GCAATCACCT AACTCCACT ACTCAGCCAT ATTCCCTGAG TCAGGGGGC CCCGGGGGGA ACTCAGCCAT ATTCCCTGAG TCAGGGGGC CCCGGGGGGA ACTCAGCCAT ATTCCCTGAG TCAGGGGGC CCCGGGGGGAC ACTCACTCCT AGGGATGCCA TCAGGGGGCC CCATACTCCAT ATATTCCACC AGGATGCAC TCTTTGAGGG CAACAAGATG TCATCCTCCA AGGCAACGAT TCTTCCTCAT GTCTCAGCA GAGCCAC CGTGCCAGAC GTGGGGACC AGGAGGGCAC TCGGGACCCG CAGGAGGATT GAATGGTTCA AGGGAATCGT AGGAGGGCT TCAGGTATCC TGGATGCTG AGGCAACGAT AGACAGAGGA TACGTGAAACC CAGGGCAC GTGGGAATCGGT AACCAGAGGA TACGTGAAACC CACAGGGCAC TTCAATCACT AAACCAGAGGA TACGTGAACC CACAGGGCAC CTTGAATCAT TCCAGACAAG GCAAACGTT CAGGAGCAC TCTATCTCT TAAAAAGGAT GTTCACCCTA AAAGCAACTT TGCACGCCTC GAAAGGCCC CTTGAATCAT TCCAGACAAG GCAAACGTT TGCACGCCTC GAAAGGCCC TTTAATCATA ATCTGCCATG AACCTTAAGCG GCAGTGGG AGAGGTATG TCAGAGCACA GTTGCAGCACAG GCAAACGTT TGCACGCCTC GAAAGGCCC TTTAATCCTT TAAAAAGGAT TTCATCCCTT AAAGCAACGAG AGATGGTCAC GAAAGACGC TGCTGAAGCAC TCATCGCTCA AAGCACAGG AGATGGTCAC GAATACGTT AGAGCACAT CAGAGAACACG GAAAACCCTT GGATCCACC CAGAGGACA GGATTGCAGA AAGCTCCAT GAATGGAAG GGTTCACCC CAGAGACACAC GAAACCACA AAGGACTCC TAATGGGAAG GGTTCACCC CAGAGGACAC GAGTTGCACAC AAAGGAAGGTT GGATCCACC CAGAGGACAC CACAAGACACAC AAAGGAAGGTT CAGATGTCC CAGAGACATC AAACCACACA AAGGAGAGGTT CAGACATTC TAAACCTTCAACCTT AAATGGAATCC GAAAACAAGG CTACCACACACACACACACACACACACACACACACACAC	ŀ	,	GATAATACCC	GAGGAGGTCC	GGGCCAACTT	TGAATACTAC	ATCGACTGGC
GGCATATTAC ACCATCGCAC GCCATATTC TAGATGATC AGGAACCTTT TGGACGATGA GTCTCTTTGAT GCCATATTC TAGATGATC AGGAACCTTT GAGGATCTCA GGGAGGAAT CCGGTATGC TAGATGATC AGGAACCTTT ACTCCTCAAAG GACCTCATAG GCAATCACCT ACTGGAGGCT ACTCACCCT ATCCCTGAG TCAGGGTGGC CCCGGGGGGC TGTGTTCTCAGA GGATGCCA TTCTTCAGAG CAACAACAGAT TCATCCTCCA AGGCAACGT CATACTCCTC AGGGATGCCA TCAGGAAGCA CGGTGCAGAC GTGGACACGT TCTTCCTCAT GTCCTCACCA GAGCATGCC CAGGAGGAT CAGGATGCA ACTGAGGGTA TCAGGGACCC CAGGAGGATT CAGGAACCA ACTGAAGAC CACAACGAGAT CACACCTGAAGCA CCCCACACGAGGACACTTCA ACTGAGAGCA CACACAGGGCA ACCTCACCA ACCTGAAGC CACACGAGGAT TCACACCAGAGA TACACTGAAACC CACAAGGGCC CTGAATCAT TCCAGCACAGA AACCACAGAGGA TACGTGAAGC CACAAGGGCC CTGAATCAT TCCAGCACAGA AACCACAGAGGACA CACAAGGGCC CTGAATCAT TCCAGCACAGA AACCACAGAGGACA CACAAGGGCC CTGAATCAT TCCAGCACAGA AACCACTCACCA GCCTGACTCT CACACCACA AAAGCACTCCACC GCACACGTT CATACCCTA CACACCACAGAGACACTC CACACGAGGACAG CACACGACACAGA AACCACCACAGAGCACACACA			TCAATGACTG	GGCATGTTCA	AGGAGGATAG	GCCTTGGAAC	AAGGCTGCCC
GGCATATTAC ACCATCGCAC GCCATATTC TAGATGATC AGGAACCTTT TGGACGATGA GTCTCTTTGAT GCCATATTC TAGATGATC AGGAACCTTT GAGGATCTCA GGGAGGAAT CCGGTATGC TAGATGATC AGGAACCTTT ACTCCTCAAAG GACCTCATAG GCAATCACCT ACTGGAGGCT ACTCACCCT ATCCCTGAG TCAGGGTGGC CCCGGGGGGC TGTGTTCTCAGA GGATGCCA TTCTTCAGAG CAACAACAGAT TCATCCTCCA AGGCAACGT CATACTCCTC AGGGATGCCA TCAGGAAGCA CGGTGCAGAC GTGGACACGT TCTTCCTCAT GTCCTCACCA GAGCATGCC CAGGAGGAT CAGGATGCA ACTGAGGGTA TCAGGGACCC CAGGAGGATT CAGGAACCA ACTGAAGAC CACAACGAGAT CACACCTGAAGCA CCCCACACGAGGACACTTCA ACTGAGAGCA CACACAGGGCA ACCTCACCA ACCTGAAGC CACACGAGGAT TCACACCAGAGA TACACTGAAACC CACAAGGGCC CTGAATCAT TCCAGCACAGA AACCACAGAGGA TACGTGAAGC CACAAGGGCC CTGAATCAT TCCAGCACAGA AACCACAGAGGACA CACAAGGGCC CTGAATCAT TCCAGCACAGA AACCACAGAGGACA CACAAGGGCC CTGAATCAT TCCAGCACAGA AACCACTCACCA GCCTGACTCT CACACCACA AAAGCACTCCACC GCACACGTT CATACCCTA CACACCACAGAGACACTC CACACGAGGACAG CACACGACACAGA AACCACCACAGAGCACACACA	j		TGGGATGAGA	GGTGGATCAT	CGAACCCCTC	ACAGACTCAA	CAATCTACAT
TGGACGATGA GTTCTTTGAT GCCATATTCC TAGATGATTCA AGGACCTTT GAGGATCTCA GGGAGGATT CCGGTACTGG TACCCCCTTG ACTGGAGGCT CTCTGCAAAG GACCTCATAG GCAACACACT TACCCCCTTG ACTGGAGGCT GCTATAGGCC TTCTTGAGGG CAACAGATT CATCCACC AGGACCGC CATACTCCTA GGGATGCCA TGAGGAGAGCA CGGTGCAGAC GTGGTGCGGC CATACTCCTA GGGATGCCA TGAGGAGAGCA CGGTGCAGAC GTGGTGCGGC AGGAGGGTTA TCGGGACCG CAGGAGGATT GAATGGTTCA CTGAGGAGGACA AGGAGGGTT CTGAGGAGCC TGGAGAGAGCA GAGCATTCGA GGGATGCTA AACCAGAGGA TACGTGAAAGC TCGAGAGGAC GTGGAGGTTA AACCAGAGGA TACGTGAAAGC TCGAGAGGCC CTTGAATCAT TCCAGCACAG AACCTCAGC AACCTTC TGGAGGCCC CAGAGGGCC CTTGAATCAT TCCAGCACAG AACCTCAGG GAACGTT CTGCACGCCT GATAAGGCC CTTGAATCAT TCCAGCACAG GCAAACGTTC TGCACGCCT GATAAGGCC CTTGAATCAT TCCAGCACAG GCAAACGTTC TGCACGCCT GATAAGGCC AAGAGGCC TTTCATCCCTT AACCTTGAGG GAACTCCCC CGAGAGGAC TCATACCCTA AAGCTCCATG GCCTGACTTC TCAGATGAT TCCAGCACAG GAACGTTC GCCTGACTTC TCAGATGATC TCAGCCACAGAGGCC AAGACGTCCACAAAGGACATC GCCTGACTTC TCAGATGAT ACCTTACCCTA AAGCACAGT GCCTGACTTC TCAGATGATC TCAGACCACG AACGTCCAT GCCTGACTTC TCAGATGATC TCAGCCCACAAAGGCC AAGACGTTC GCTGACTTC TCAGATGATC TCATACCCTA AAGCACAGT GCCTGACTTC TCAGATGATC TCATACCCTA AAGCACAGT GCCTGACTTC TCAGATGATC TCATACCCTA AATGGAAACTCTT GACTCCCC CGGAGAGGCC CAGAGGACATTC AGGAAATCCAT AATGGAACTCTT GAATCCACC CCGGAGAGGCT CACAATATAC ACCTCACCAA AATGGACATTC GGTGAATCACC CCGGAGAGGAC CACAATATAC ACCTCACCAA AATGGACATTC GGTGAATGAC GAGTACACCC AGAGGACATTC AGAGACATCACAAGCACAACAGCACATCCACAATATAC ACCTCACCAAACAGCACACACACACACACACACACACACA			GGCATATTAC	ACCATCGCAC	ACCGCCTCAG	GGAGATGGAT	GCCGGGGAGA
GAGGATCTCA GGGAGGATT CTCTGCAAAG GACTCATAG GCAATCACT GCAATTCCAT ATATCCACC ACTCACAGCAT ATTCCCTCA ACTCACCCAT ATTCCCTCAG ACTCACCCAT ATTCCCTCAG ACTCACCCAT ATTCCCTCAG GGTATGGGC CATACTCCT AGGGATGCC CATACTCCT AGGGATGCC CATACTCCT AGGGATGCC CATCACCCT AGGGATGCC CATCACCCT AGGGATCC CATCACCCT AGGGATCC CATCACCCT AGGGATCC CATCACCCT AGGGATCC CATCACCCC CAGGAGGAT CAGGAGGAT CAGGAGGAT CAGGAGGAT CAGGAGGAT CAGGAGGAT CAGGAGGAT CAGGAGGAT CACCTGAAGC CACAGAGGC CACAGGCC CACAGGGC CACAGGCC CACAGGGC CACAGG	ł		TCCACCATCA	CUTCUTUCAT	GCCATATTCC	TAGATGATTC	AGGAACCTTT
CTCTGCAAAG ACTCAGCCAT ACTCAGCCAT ACTCAGCCAT ACTCAGCCAT ACTCAGCCAT GGTATGAGGC GGTATGAGGCC CATACTCCTGA GGTATGAGGC CATACTCCTC GGTATCTCCTCAT CATCTCCTCAT CATCAGCCAT CATCCTCATC CATACTCCTC AGGGATGCCA CATCACCCCC CAGAGAGCAT CATCCCCCA CAGCCATGCC AGGCAATGCC CAGAGGCAT CAGGAGGCAT CAGGAGGATC CAGCCATGC AGGCAATGCC CAGAGGAT CAGCGAAGCAT CAGCCATGC AGGCAATGCC CAGAGGAT CAGCGAAGCAT CAGCGAAGCAT CACCGAGAGCAT CACCGAGAGCAT CACCGAGAGCAT CACCGAGAGCAT CACCGAGAGCAT CACCGAGAGCAT CACCAGAGGAC CACACGGCC CAGAGGGCC CAGAGGGCC CAGAGGCCAT CACACGACGA CACCTTCAGCC CACACGGCC CAGAGGCCC CAGAGGCCC CAGAGGCCC CAGAGGCCC CAGAGGCCC CAGAGGCCC CAGAGGCCC CAGAGGCCC CACACGCCC CACCCCC CACCCCCC CAGAGGCCC CAGAGGCCC CAGAGGCCC CACCCCCC CACCCCCC CACCCCCC CACCCCCC CACCCCCC			CACCAMCMCA	CCCACCAATT	CCGCTACTGG	TACCCCCTTG	ACTGGAGGCT
ACTCAGCCAT ATTCCTGAG TCAGGGTGGC CCGGGGGGC TGTGTTTTTTGGGG CATACTCCTC AGGGCAACGT TCATCCTCCA AGGGCAACGT TCATCCTCCA AGGGCAACGT TCTTCCTCAT AGGGAACGA CGTGCAGACG CATCTCCAGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	ļ		GAGGATCICA	GGGAGGAATI	CCGGIACIGG	CACATTCCAT	ΔͲΔͲͲϹϹΑϹϹ
GGTATGGGCC CATACTCCTCCT CATACTCCTCCT CATACTCCTCCT AGGGATGCCA CTCTCAGCA AGTGAGGTCC AGGGATGCCA CTCGAGAAGCA CGGTGCAGAGC AGGCATTCGA AGAGAGGGTC AGAGAGGGTC CTCCAGCTGA AACCAGAGGA AACCAGAGGA CTCAGAGTT CTCAGAGAGC CTCCAGCTGA AACCAGAGGA CTCAGAGGCC CTCAGCTGA AACCAGAGGA CACCTGAAAGC CTCAAGGCC CTCAGAGGC CTCAGAGGC CTCAGAGGC CTCAGAGGC CTCAGAGGC CTCAGAGGC CTCAGAGGC CTCAGAGGC CTCAGAGGC CTCAGAGGC CTCAGAGGC CTCAGAGGC CTCAGAGGC CTGAAAGGC CTGAAAGGC CTGAAAGGC CTGAAAGGC CTGAAAGGC CTGAAAGGC CTGAAAGGC CTGAAAGGC CTGAAAGGC CACTGCAGG CACAGGGCC CTGAAAGGTTA ACCTTAAGGC CACTGCTGG CGCAGACCT CACTGCTGG CACAGGTCC CACTGCTGG CGCAGACCT CACTGCTGG CGCTGACTTC CTCAGAGGAC CACAGGGCC CTGAATTC CTCAGAGGAC CACAGGGCC CTGAATCCT CACTGCTGG CGCTGACTTC CTCAGAGGAC CACAGGGCC CTGAATCC CACAGGCC CGCTGACTTC CTCAGAGGAC CACAGGGCA CTCAGCCAG CACAGGTAGC CACAGGGCA CTCACCCA AATGGAACCAC CGGAACCACC CGGAGAGGGT CACAGAGGACA CACAGGGCA CCCCC CGGAGAGGGT CACACATAC CACAGGCAGC CCCGAAACCACC CGGAGAGGGT CACACATAC CACAGGCAGC CCCGAAACCACC CGGAGAGGGT CACACACACA AATGGACCT CACAGGAGGT CCCGAAACCACC CGGAGAGGGT CACACACACA CACAGGGTAG CACAGGGTAG CACAGGGTAG CACAGGGTAC CACACACACA CACAGGGTAC CACAGGGTAC CACAGGGTAC CACAGGGTAC CACACACACACAC CACAGGGTAC CACACACACACACACACACACACACACACACACACA			CTCTGCAAAG	GACCTCATAG	GCAATCACCT	GACATICCAT	TIMITCONO CONTO
CATACTCTG AGGATGCCA TCGAGAAGCA GGTGCAGAC CTGGAGGGAC AGTCTCTCATCTCTCATCTCTCATCTCTCATCTCTCAGCA GAGCCATGGC AGACCATTGC AGAGCATTCA GGGAATTCAG AGAGAGAGGACT TCGAGACACC CAGAGGATT GAGAGAGGAT AGAGAGAGGACT CTCAGAGAGC TTCATTGCAAC GGGAATTCAG AACCAGAGGA TACCTGAAAGC TCTATTCACT TAAAAAAGGAT TCCAGACAAG AAACCACAGAGC ACCAGAGGCC CTTATCACT TCCAGACAAG AACCACAGAGCT CTAGACACAC ACCAGAGGCA TCCATCACACACACACACC CACAAGGCC CTTATCACT TCAAAAAAGGAT GTTGACCACT ACCTTAAGCCT TCACACACG GAGATATGG AAGGTCCAT GCACACGC CACAAGGCC CTCACTCCAT TCATACCCTA ACCTTCAGCACAG GAGATGTGG AAACGTCCAT GCACACCC CACAGAGCAC GAGAGACAC ACCTCCACT GAAACACCT CACTGCTGAG AAGGTCCAC GAAACACCT CACTGCAGAG AAGGTCCAC GAAACACCT CACTGCAGAG AAGGTCCAC GAATACCACC CACAGAGAGCAC GAACACACC CACAGAGAGCAC GAACACACC CACAGAGAGCAC GAACACACC CACAGAGAGAC ACCTCACACA AATGGAAATG GAGACATCT CACAGAAGAC AATGGAACACC TAATGGAACA GAATACCAT ACCACAA AATGGAACATC GAATCACCA AATGGAACAC GAATACCAC AATGGAACAC GAATCACAC AATGGAACAC GAAGACATTC CACAGAGTTC CACAGAGACAC AATGGAACAC GAACACACAC AATGGAACAC GAACACACAC AATGGAACAC CACAGAGACC AATGACAC AATGGAACAC CACAGAGACC AATGACAC AATGGAACAC CACAGAGAC CACAGAGACC AATGACACAC AATGGAACAC CACAGAGACC AACAGACCACAC AATGACACAC CACAGAGACC AACAGACCACAC CACAGAGACC AACAGACACACAC			ACTCAGCCAT	ATTCCCTGAG	TCAGGGTGGC		1G1GGICIII
TCTTCCTCAT GTCCTCAGCA GAGCCATGGC AGGACTTTGA CTGGAGGAGA AGTGAGGTCA TCGGGACCCG CAGGAGGATT GAATGGTTCA GGGAATTCGG AGAGGGTC TCAGGTATCC TGGATGGTAG GCCAGTCCTC AGTGAGGTTA ACCAGAGGA TACGTGAAGC TCTATCTCCT TGAATCAT TCCAGACAG AACCAGAGGA TACGTGAGC CACAAGGGCC CTTGAATCAT TCCAGACAG AACCATAGCG TGTTGAGGGT AGAGTTGATG ACCTTAAGACG TCTATCTCCT TAAAAAGGAT TTCATCCTT ACCTTCAGG GCAAACGTTC TCCACCCCTG GATTAGTCT TCATCCCTA ACCTTCAGG GAGATTCGG GAGATGATG TGTGAGGGT ATGAGGATAA ATCTGTCCTT CACTCCATG GAGATCCTC CACTGATTG TCAGAGCAGG TTGATGAGG ATGACCTTA ACCTTAGAGCT GAGAGATCATT GTTGCAGAGG AAGATCCTT GAGATGATG CAGAGGAGAGG	l		GGTATGGGCC	TTCTTGAGGG	CAACAAGATG	TCATCCTCCA	AGGGCAACGT
AGTGAGGTCA TCGGGACCCG CAGGAGGATT GAATGGTTCA GGGAATTCGG AGAGAGGGTC TCAGCTATCC TGGATGGTAG GCCAGTCCTC ACTCAGCTGA ACCTGAAAGC TTCATTGGAA GGTGGATGAT TCCAGCAGAGGAACC TCTATTCCT TAAAAAGGAA TCCGTCAAGC CACAAGGGCC AAAAGGCAGTT CAGGAGGACC CACAAGGGCC CTTGAATCAT TCCAGACAAG TACCTTAAACGA TACCTTAAAGCG TGTTGAGGGT AGAGTTGATG ATGAGGTTAA ATCTGTCCTT TAAAAAGGAA ATCTGTCCTT TGCACGCCTG GATAAGCCTC ATGAGGTTAA ATCTGTCCTT TCACATGCTAG GGTGATGATG ATGAGGTTAA ATCTGTCCTT TCACATGCTAGA AAGCTCCAAG GAGAGTTTCG GAGAGTTCGA GAGAGACTATC CACAGAGAGCAG GAGAGTTCGA GAGAGACCATC CACAGAGAGCAG GAGAGTCCAAG GAATCCATC CAGAGAGCAG GAGAGACCATC AAAGCACACC CGGAGAGGGT CACAGAGACATC AAACCACAAAACAATC AATGGAAATG GGATGTCCAA GAGAGACATTC CACACATATAC ACCTCACCAA AATGGAAATG GGATGTCCAA GAGAGACATTC CACACATATAC ACCTCACCAA AATGGAAATG GGATGTCCAA GAGAGACATTC CACACAGATTC CACACAGATTC CACACAGATTC CACACATATAC ACCTCACCAA AATGGAAATG GGATGTCCAG GAGGGACGC CACACAGGACGAG GGTTCCACC CACACAGATAC ACCTCACCAA AAGCAGACAGAC GAGAGACAGC CACACAGAACAGA GGATACACCG TACACAGAACAGA			CATACTCCTG	AGGGATGCCA	TCGAGAAGCA	CGGTGCAGAC	GTGGTGCGGC
AGAGAGGGTC TCAGGTATCC TGGATGGTAG GCCAGTCCT AGTGAGGTTAGAT CTCCAGCTGA ACCTGAAAGC TTCATTGGAA GGTGATGAT GGGTCAGCTG AACCAGAGGA TACGTGAAGC CACAAGGGCC CTTGAATCAT TCCAGACAGA AACGAGGATT CAGGAGGCAC TCTATCTCCT TAAAAAGGAT ACCTTAAGCG TGATGAGGTTAA ACCTGAGAGGTAGAT GGAGAGGTTC TGCAGGCTG GAAAAGGGCT CACTGCTGAGGGT CACTGCTGAGGGT CACTGCTGAGGGT CACTGCTGAGGGT CACTGCTGAGGGT CACTGCTGAGGGT CACTGCTAGG GAGAGTGTGG AAGGTCCAT GCACGCTG GATAAGGGT CACAGAGGCAG GGATGTCCAG GAGAGCTCAG CACAGAGGCAG GGATGTCCAG GAGAGTCCTA AATGGAAAGG GGATCCACC CGGAGAGGGT CACAGAGAGCAG GGATGTCCAG GAAGAGCAG GGATGTCCAG GAAGAGCAG GAAGAGCAG AAAACCAGAT AATGGAAATG GGATCCACC CGGAGAGGGT CACAGAGGAG AAAACCAGAT AATGGGAAAG GGATGTCAA AGGGAAAGAAG GGATTCAGCT GAAGAGGAG CACAGAGGAG AAAACCAGAT AATGGGAAGG GGATGCCAC CACAATACA AACCAACAGA AATGGGAATG CAGAGGGTAGG GAAGAGGAGGT CACAGAGGAGGT ACACAGA GAAACAAGAG CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGG TACACAGA CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGG TACACAGA CACAGAGGAGGT CACAGAGGAGG TACACAGA CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGG TACACAGA CACAGAGGAGGT CACAGAGGAGG TACACAGA CACAGAGGAGG TACACAGA CACAGAGGAGG TACACAGA CACAGAGGAGGT CACAGAGGAGG TACACAGAGG TACACAGA CACAGAGGAGGT CACAGAGGAGG TACACAGA CACAGAGG TACACAGA CACAGAGGAGG TACACAGA CACAGAGGAGGT CACAGAGGAGG TACACAGA CACAGAGAGG TACACAGA CACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG			TCTTCCTCAT	GTCCTCAGCA	GAGCCATGGC	AGGACTTTGA	CTGGAGGGAG
AGAGAGGGTC TCAGGTATCC TGGATGGTAG GCCAGTCCT AGTGAGGTTAGAT CTCCAGCTGA ACCTGAAAGC TTCATTGGAA GGTGATGAT GGGTCAGCTG AACCAGAGGA TACGTGAAGC CACAAGGGCC CTTGAATCAT TCCAGACAGA AACGAGGATT CAGGAGGCAC TCTATCTCCT TAAAAAGGAT ACCTTAAGCG TGATGAGGTTAA ACCTGAGAGGTAGAT GGAGAGGTTC TGCAGGCTG GAAAAGGGCT CACTGCTGAGGGT CACTGCTGAGGGT CACTGCTGAGGGT CACTGCTGAGGGT CACTGCTGAGGGT CACTGCTGAGGGT CACTGCTAGG GAGAGTGTGG AAGGTCCAT GCACGCTG GATAAGGGT CACAGAGGCAG GGATGTCCAG GAGAGCTCAG CACAGAGGCAG GGATGTCCAG GAGAGTCCTA AATGGAAAGG GGATCCACC CGGAGAGGGT CACAGAGAGCAG GGATGTCCAG GAAGAGCAG GGATGTCCAG GAAGAGCAG GAAGAGCAG AAAACCAGAT AATGGAAATG GGATCCACC CGGAGAGGGT CACAGAGGAG AAAACCAGAT AATGGGAAAG GGATGTCAA AGGGAAAGAAG GGATTCAGCT GAAGAGGAG CACAGAGGAG AAAACCAGAT AATGGGAAGG GGATGCCAC CACAATACA AACCAACAGA AATGGGAATG CAGAGGGTAGG GAAGAGGAGGT CACAGAGGAGGT ACACAGA GAAACAAGAG CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGG TACACAGA CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGG TACACAGA CACAGAGGAGGT CACAGAGGAGG TACACAGA CACAGAGGAGGT CACAGAGGAGGT CACAGAGGAGG TACACAGA CACAGAGGAGGT CACAGAGGAGG TACACAGA CACAGAGGAGG TACACAGA CACAGAGGAGG TACACAGA CACAGAGGAGGT CACAGAGGAGG TACACAGAGG TACACAGA CACAGAGGAGGT CACAGAGGAGG TACACAGA CACAGAGG TACACAGA CACAGAGGAGG TACACAGA CACAGAGGAGGT CACAGAGGAGG TACACAGA CACAGAGAGG TACACAGA CACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	ļ		AGTGAGGTCA	TCGGGACCCG	CAGGAGGATT	GAATGGTTCA	GGGAATTCGG
CTCCAGCTGA ACCTGAAAGC TTCATTGGAA GGTGGATGAT GGGTCACGTG AACCAGAGGA TACGTGAAAGC CACAAGGGCC CTTGAATCAT TCCAGACAAG AAAAGGCAGTT CAGAGAGCA TCTATCTCT TAAAAAAGGAT GTTGACCACT ACCTTAAGCG TGTTGAGGGT AGAGTTGATG ATGAGCTCAT GCAAAACGTTC TGCACCCCTG GATAAGGCTC ATGGCTCAT TCATACCCTA CACTGCTGAG GAGATGTGGG AGAGTATGG TGGTGAGGGT TTTGTAGCAG AAGCTCCATG GCCTGACTTC TCAGATGATG CAGAGAGCACA GGAATCCATG GAAGATCCATG GCTGACTTC TCAGATGATG CAGAGAGCACA GGAATCCATG GAAGATCCTT GGATCCACC CGGAGAGGGGT CACACATATAC ACCTCACCAA AATGGAAATG GAATCCACC CGGAGAGGGGT CACACATATAC ACCTCACCAA AATGGAAATG GAATCCACC CGGAGAGGGGT CACACATATAC ACCTCACCAA AATGGAAATG GAATCCACC CGGAGAGGGT CACACATATAC ACCTCACCAA AATGGAAATG GAATCCACC CGGAGAGGGT CACACATATAC ACCTCACCAA AATGGAAATG GAATCACCCT CAGAGGGTT CAGAGGACATC ATGATAACAT CAGAGGTTAC CAGAGGTTG CAGAGGCATC CATGATAACAT CAGAGGTTAC CAGAGGTTGG AGCCACGCT TACTCATGGA TACACTGATA CAGAGGTTAC CAGAGGTTGG AGCCACGCT TACTCATGGA TACACTGATA CAGAGGTTAC CAGAGGTTGG AGCCACGCT TACTCATGAC CATCACCAA AAACAAAGG CTGTGAATGC CAGTCACCC CACACACAAACAGA CAAACAAAG CTTGGAATACAC CATCAGGGA CAGAGCATCC TACACTAGAG CACACAGA AAACAAAAAAAGG CTGTGAATGC CAGTCACCCC TATACCCTTA ATGA CATACACTTGA ATGA SEQ ID: 15 Archaeog lobus fulgidu	Ī		AGAGAGGGTC	TCAGGTATCC	TGGATGGTAG	GCCAGTCCTC	AGTGAGGTTA
AACCAGAGGA TACGTGAAGC CACAAGGGCC CTTGAATCAT TCCAGACAGA AAAGGCAGTT CAGAGAGCAC TCTATCTCCT TAAAAAAGGAT GTTGACCACT ACCTTAAGCG TGTTGAGGGT AGAGTTGATG ATGAGGTTAA ATCTGTCCTT GCAAACGTTC TGCACGCCTG GATAAGGGCT ATGGCTACAT TCCATACCCTA CACTGCTGAG GAGATCTGG AGAGTATGG TGGTGAGGGT TTTGTAGCAG AAGCTCCATG GCCTGACTTC TCAGATGATG CAGAGAGCACG GGATGTCCAG GATGCCACACC GGAGAGCACT CAGAGAGCACG GGATGTCCAG GAAGATCCTT GGATCACCC CGGAGAGGGT CACACATATAC ACCTCACCAA AATGGAAATG GGATCTCAACCC CGGAGAGGGT CACACATATAC ACCTCACCAA AATGGACATC TAATGGGAAG GGTTTCAGCT GAGGGCATCC ATGATAACAT CAGAGGTTAC GCTGAATTG TAAGGAGGAT CATCAGGGA CAGAGGTTAG CAGAGGTTAC CAGAGGTTGG AGCGCT TACTCAGCAA ATGGACATC CACACACC CAGAGGGT CATCATGGA AAACAAGAT CAGAGGTTAC CAGAGGTTGG AGCCACCCA CAGAGGTAGG CAGAGCATC CACATATAC CACACACACACACACACACACACAC			CTCCACCTCA	ACCTGAAAGC	ттсаттссаа	GGTGGATGAT	GGGTCAGCTG
AAAGGCAGTT ACCTTAAGCG GCAAACGTTC CACTGCTGAG GAATTCTCT CACTGTGAG AAGCTCCAT GCACACGTC CACTGCTGAG AAGCTCCAT CACTGCAGAGG CACTGCAGAGG CACTGCAGAGG CACTGCAGAGG CACTGCAGAGG CACTGCAGAGG CACTGCAGAGG CACTGCAGAGG CACTGCAGAGG CACTGCAGAGG CACAGAGGAGG CACAGAGGAGG CACAGAGGAGG CACAGAGGAGG CACAGAGGAGG CACAGAGGAGG CACAGAGGAGG CACAGAGGAGG CACAGAGGAGG CACAGAGGAGG CACAGAGGAGG CACAGAGGAG CACAGAGGAG CACAGAGCAG CACAGAGGAG CACAGAGAGCAG CACAGAGAGCAG CACAGAGAGCAG CACAGAGACATC CACAGAGGAGG CACAGAGGAG CACAGAGAGCAG CACAGAGAGCAG CACAGAGAGCAG CACAGAGAGCAG CACAGAGGAGC CACAGAGGAG CACAGAGAGCAG CACAGAGGAGC CACACAGAACAGA CACACAGA CACAGAGGAGC CACAGAGGAG CACAGAGGAG CACACAGAACAGA CACACAGA CACACAGA CACACAGA CACACAGA CACACAGA CACACAGA CACACAGAACACAC CACAGAGAGCAC CACACACA			A D CC D CC C C	MACCIGAMAGC	CACAAGGGCC	CTTGAATCÁT	TCCAGACAAG
ACCTTAAGCG GCAAACGTTC GCACGCCTG GCATAAGGCT CACTGCTGAG AAGCTCCATC CACTGCTGAG AAGCTCCATC CACTGCTGAG AAGCTCCATC CACTGCTGAG CACTGAGAGG CACTGACTCC CACTGACTCC CACTGACTCC CACTGACTCC CACTGACTCC CACATATAC CACTGAGAG CACTCACCC CAGAGGAGCAG CAGAGACCACC CAGAGAGCCAG CAGAGATCCACC CAGAGAGCCAG CAGAGACCACC CAGAGGACCACC CAGAGGACCAC CAGAGACCACC CAGAGGACCAC CAGAGGACCAC CAGAGGACCAC CAGAGGACCAC CAGAGGACC CACACATATAC CACACAA AAACACAA CACCCACACAA CACCCACACAA CACCCACACACAC CAGAGGACCC CAGAGGACCC CACACATATAC CAGAGGACCC CACACACACA CACCCACACACAC CAGAGGACCC CACACACA			AACCAGAGGA	TACGIGAAGC	TOTAL TIPE TOTAL	TAAAAACCAT	CTTCACCACT
GCAAACGTTC CACTGCTGAG CACTGCTGAG CACTGCTGAG CACTGCATG CACTGCATG CACTGCAGG CACTGCATG CACTGCAGG CACTGCATG CACTGCAGG CACTGCATC CACTGCAGG CACTGCACTC CACAGGAGCAC CACAGAGCAC CACAGAGCAC CACAGAGCAC CACAGAGCAC CACAGAGCAC CACACATATAC CACAGCACC CACACATATAC CACAGCACC CACACATATAC CACAGCACC CACACATATAC CACACCAC CACACATATAC CACACCAC CACACATATAC CACACCAC CACACATATAC CACACCAC CACACATATAC CACACCAC CACACATATAC CACACCAC CACACATATAC CACACCAC CACACATATAC CACACCACA CACACCACC CACACATATAC CACACCAC CACACCACA CACACCACA CACACCAC			AAAGGCAGTT	CAGGAGGCAC	1C1A1C1CC1	አመር አርርጣጥ አ	ያብረብርብርርብርርብብ የተ
CACTGCTGAG AAGCTCCATG CGTTGCACAGG CGTTGCACAGGG GAAGATCCT GAAGATCCTT CGAAGAGGGGG CGAAGACCAG GAAGACCAG CCACATATAC AAGGAAATCA AAGGAACATC CACACATATAC AAGGAACAG CCACATATAC AAGGAACAG CCACATATAC AAGGAACAG CCACATATAC AAGGAACAG CCACATATAC AAGGAACAG CCACATATAC CACACAGGTAG AAAACAGAG CCTGAATTG CAGAGGTTAC GAAGAGAGCAG GATTCAGCCT GAAGGAGGAG CAGAGGTAG CACACAGAGTAG AAAACAGAG CATCAGGCA CACAGGGTAC CAGAGGTAC CAGAGGTAC CAGAGGTAC CAGAGGTAC CAGAGGTAC CAGAGGTAC CAGAGGTAC CAGAGGTAC CAGAGGTAC CACACAGAC CACACAGAC CACACAGAC CACACAGAC CACACAGAC CACACAGAC CACACAGC CACACACA			ACCTTAAGCG	TGTTGAGGGT	AGAGTTGATG	ATGAGGITAA	MICIGICCII
AAGCTCCATG GTTGCAGAGG GAAGATCCTC GTTGCAGAGG GAAGATCCTC GAAGATCCTC GAAGATCCTC GAAGATCCACCC GAAGATCCACCC GAAGACACTC AATGGAATC AATGGAATC AATGGAATC AATGGACTCA AATGGACACCC GGATCCACCC AATGGCTCCA AATGGACTCA AATGGACTCA GAAGGAGGTT GCAGAGGTTC GAAGGAGGTT CAGAGGTTAC GAAGGAGGTT CAGAGGTTAC CAGAGGTTAC CAGAGGTTAC CAGAGGTTC CACACAA AGCCAGGGT CACTCATGAA CCTCTGAT CACACCA CACACACA CCTCCACTA ACCCAGGAT CACCAGGAC CTTGGTAAAC CCACACACA CCACCACCC CGTTCCCCT CAGACCAG CTGTGAATC CACAGGACC CTTGGTAAAC CCACACACA CCACACACC CGTTCCCCT CACACACAC CGTTCCCCT CACACACAC CGTTCCCCT CACACACCA CCACACACC CGTTCCCCT CACACACAC CGTTCCCCT CACACACAC CGTTCCCCT CACACACAC CGTTCCCCT CACACACAC CGTTCCCCT CACACACAC CGCTCACT CACACAC CACACACAC CGAGGTAC CCACATATAC CACACACAC CAGAGGTAC CACACACAC CTTGGTAAAC CCACACACAC CGTTCCCCT CACACACAC CGAGGTC CACACACAC CACACACC CGTCCCCT CACACACC CACACACC CTTGTAACC CACACACC CTTGTAACC CACACACC CTTGTAACC CACACACC CTTGTAACC CACACACC CTTGTAACC CACACACC CTTCCCCT CACACACC CTTGTAAC CACACACC CTTCCCTC CACACACC CTTGTAACC CACACACC CTTCCCTC CACACACAC CTTCCCTC CACACACC CTTGTAAT CACCACAC CTTCATCAC CACACACC CTTGTAACC CACACACC CTTGTAACC CACACACC CTTGTAACC CACACACC CTTGTAACC CACACACC CTTGTAACC CACACACC CTTGTAACC CACACACC CTTGTAACC CACACACC CTCACAC CTCCCTC CACACACC CTCCCT CACACACC CTCCCTC CACACAC CTCCCT CACACACC CTCCACAC CTCCCTC CACACACC CTCCCC CACACACC CTCCCC CACACACC			GCAAACGTTC	TGCACGCCTG	GATAAGGCTC	ATGGCTCCAT	TCATACCCIA
GTTGCAGAGG GAAGATCCTT GAAGATCCTT AATGGAAATG AATGGAAATG AATGGAAATG AATGGAAATG AATGGAAATG AATGGAAATG AATGGAAATG AATGGAAATG AATGGAAGT CAGAGGTTAC GAAGGTTAC GAAGGAGGT CAGAGGTTAC GAAGGAGGT CAGAGGTTAC GAAGGAGGT CAGAGGTTAC CATCAGGGAC CTATGAACCA AGCCAGGGTT GTGATACACA AGCCAGCCA AAGCCAGC CAGTCCACC CATCAGGGAC CTATGAACCA CAGAGGTACC CATCAGGGAC CTACTAGAC CAGAGATACAC CATCAGGGAC CTACTAGAC CAGAGATACAC CATCAGGGAC CTACTAGAC CAGAGATACAC CATCAGGGAC CTACTAGAC CAGAGTACAC CATCAGGGAC CTACTAGGAC CAGCATACAC CATCAGGGAC CTACTAGGAC CATCAGGGAC CTACTAGGAC CACTCATGAA TACCTTGA AGCCAGCCA AAGCCAGC CAGAGTACC CATCAGGGAC CTACTAGGAC CTACTAGGAC CTACTAGGAC CTACTAGGAC CTACTAGGAC CTACTAGGAC CTACTAGGAC CTACTAGAC CATCAGGGAC CTACTAGGAC CTACTCATGA AGCCAGCT AAGCCAGC AAACAACA CAGCAGGTTAC CTACTAGGAC CTACTGAT AAGCCAGC CTGTGATAC CTACTAGGAC CTACTAGGAC CTACTAGGAC CTACTGGA AGCCAGGAT CTACTGAT CATCACTGA AAGCAACAAG CTACACAA AGCCAGCA ACACAGA CTACACAC AGCAGATACA CATCAGGAC CTACACAC AGCAGATACA CATCAGGAC CTACACA AGCCAGCA AAGCAACAAG CTACACAC AGCAGATACA CATCAGGAC CTACACAC AGCACACA CACACACA CACACACA CACACACA			CACTGCTGAG	GAGATGTGGG	AGAGGTATGG	TGGTGAGGGT	TTTGTAGCAG
GAAGATCCTT GGATCCACCC CGGAGAGGGT CCACATATAC ACCTCACCAA AATGGAAATG GGATGTGCTA AGGGTCGCAG CAGAGGTAGG AAAACTAGAT ATGGGCTCCA TAATGGGAAG GGTTTCAGCT CATCAGGGAC CAGAGGTACC CAGAGGTACAC GAAGGGTTAC CAGAGGTTAG CAGAGGATACAC CAGAGGTACACAC CAGAGGTTAC CAGAGGTTAG CAGAGGATGC CAGAGGGTT TAACATTGAAT CACATTGAAT CAGAGACACG CAGAGGAGAC CATCAGGGAC CAGAGGATACACAC GAGAGACCAGA GCCAGAGGAT CACACAGA GCCAGACCACACACG CAGACCAGA GCCAGAGACCAGA CAGACCAGA CAGACACAGA CAGACACAGA CAGACACAGA CAGACACAGA CAGACACAGA CAGACACAGA CAGACACAGA CAGACACAGA CAGACACAGA CAGACACACA CAGACACACA CAGACACACA CAGACACACA CAGACACACAC			AAGCTCCATG	GCCTGACTTC	TCAGATGATG	CAGAGAGCAG	GGATGTGCAG
GAAGATCCTT GGATCCACCC CGGAGAGGGT CCACATATAC ACCTCACCAA AATGGAAATG GGATGTGCTA AGGGTCGCAG CAGAGGTAGG AAAACTAGAT ATGGGCTCCA TAATGGGAAG GGTTTCAGCT CATCAGGGAC CAGAGGTACC CAGAGGTACAC GAAGGGTTAC CAGAGGTTAG CAGAGGATACAC CAGAGGTACACAC CAGAGGTTAC CAGAGGTTAG CAGAGGATGC CAGAGGGTT TAACATTGAAT CACATTGAAT CAGAGACACG CAGAGGAGAC CATCAGGGAC CAGAGGATACACAC GAGAGACCAGA GCCAGAGGAT CACACAGA GCCAGACCACACACG CAGACCAGA GCCAGAGACCAGA CAGACCAGA CACACAGA CAGACCAGA CAGACCAGA CAGACCAGA CAGACCAGA CAGACCAGA CAGACCAGA CAGACCAGA CAGACCAGA CAGACCAGA CAGACCAGA CAGACCAGA CAGACCAGA CACACCG CAAACCAGA CAGACCAGA CAGACCAGA CAGACCAGA CAGACACAGA CAGACACAGA CAGACACAGA CAGACACAGA CAGACACAGA CAGACACAGA CAGACACAGA CAGACACAGA CAGACACAGA CAGACACAGA CAGACACAGA CAGACACAGA CAGACACAGA CAGACACAGA CAGACACACA CAGACACACAC		•	GTTGCAGAGG	AGATGGTCCA	GAATACCGTT	AGAGACATTC	AGGAAATCAT
AATGGAAATG ATGGAAATG ATGGGCTCCA TAATGGGAAG GGTTTCAGCT GAGGGCATCC TAAGGAGGAT CATCAGGGAC GAAGGAGGTTAC CAGAGGTTAC CCAGGAC ACCAGGAC CACCAGAC CACCAGAC CACCAGAC CACCAGGAC CACCAGGAC CACCAGGAC CACCAGGAC CACCAGGAC CACCAGAC CACCAGGAC CACCAGAC CACCAGAC CACCAGAC CACCAG CACCAGGAC CACCAGAC CACCAGGAC CACCAGAC CACCAGAC CACCAGAC CACCAG CACCAGAC CACCAG CACCAG CACCAGGAC CACCAGAC CACCAG CACCAGGAC CACCAGAC CACCAG CACCAGAC CACCAGGAC CACCAGAC CACCAG CACCAGAC CACCAG CACCAGGAC CACCAGAC CACCAC CACCAGAC CACCAG CACCAC CACCAG CACCAC CACCAG CACCAC CACCAC CACCAC CACCAC CACCAC CACCAC			GAAGATCCTT	GGATCCACCC	CGGAGAGGGT	CCACATATAC	ACCTCACCAA
ATGGGCTCCA TAATGGGAAG GGTTTCAGCT GAGGGCATCC CTTGGTAAAT GAAGGAGGTT GCTGAATTTG TAAGGAGGAT CATCAGGGAC CTTGGTAAAT CAGAGGTTAC GGTGATAGAC GAGTACAGCG TACTCATGGA TGCATCTGAT TACATTGAAT CAGAGGTTG AGCCAGGGTT GTGATACACA GCAAACCAGA CTATGACCCT GAAAACAAGG CTGTGAATGC CGTTCCCCTG AAGCCAGCCA TATACCTTGA ATGA SEQ ID: 15 Archaeog lobus fulgidus leucyl trna- syntheta se (AFLRS) RS QWFLNYSNRE WKEKVLNHLE KMRIIPDYYK SEKPVICRCG TKCVVKVVRD RS QWFLNYSNRE WKEKVLNHLE KMRIIPDYYK EFFRNKIEWL RDKACARKG QWFLNYSNRE WKEKVLNHLE KMRIIPDYYK EFFRNKIEWL RDKACARKG LGTRIPWDKE WLIESLSDST IYMAYYILAK YINAGLLKAE NMTPEFLDYV YLFHHVAIFP PDKWPRAIAV NGYVSLEGKK MSKSKGPLLT MKRACVQVGAA			AATCCAAATC	GGATGTGCTA	AGGGTCGCAG	CAGAGGTAGG	AAAACTAGAT
GAAGGAGGTT CAGAGGTTAC CAGAGGTTAC CAGAGGTTAC CAGAGGTTAC CAGAGGTTAC CAGAGGTTAC CAGAGGTTAC CAGAGGTTGC CAGAGGTTGC CAGAGGTTGC CAGAGACAAGG CTGTGAATGC CTGTGAATGC CTGTGAATGC CTGTGAATGC CTGTGAATGC CTGTGAATGC CTGTGAATGC CTGTGAATCACA GCAAACCAGA CGTTCCCCTG CGTTCCCCTG CGTTCCCCTG CGTTCCCCTG CGTTCCCCTG CGTTCCCCTG CGTTCCCCTG CGTTCCCTG CGTTCCCCTG CGTTCCCCTG CGTTCCCCTG CGTTCCCCTG CGTTCCCCTG CGTTCCCTG CGTTCCCCTG GTTCCCCTG GTTCCCCTG CGTTCCCCTG CGTTCCCCTG CGTTCCCCTG CGTTCCCCTG CGTTCCCTG CGTTCCCCTG CGTTCCCCTG CGTTCCCCTG CGTTCCCCTG CGTTCCCCTG CGTTCCCTG CGTTCCTCTC CGTTCCCTG CGTTCCCTG CGTTCCCTCT CGTTCCCTCT CGTTCCCTCT CGTTCCCTCT CGTTCCCTCT CGTTCCC	•		אתכפפפיים	ቸል አጥርርርል ልር	GGTTTCAGCT	GAGGGCATCC	ATGATAACAT
CAGAGGTTAC GGTGATAGAC GAGTACAGCG TACTCATGGA TGCATCTGAT TACATTGAAT CAGAGGTTGG AGCCAGGGTT GTGATACACA CTATGACCCT GAAAACAAGG CTGTGAATGC CGTTCCCCTG AAGCCAGCCA TATACCTTGA ATGA SEQ ID: 15 Archaeog lobus fulgidus leucyl trna-syntheta se (AFLRS) RS QWFLNYSNRE WKEKVLNHLE SQKDKELDK (AFLRS) RS QWFLNYSNRE WKEKVLNHLE KMRIIPDYYK EEFRNKIEWL LGGKGEVGV AEASKLSVEL GQGRDFEY WYPVDLRSG KDLVANHLLF LGGKGEVGV AEASKLSVEL IQQIRDDFEY WYPVDLRSG KDLVANHLLF LGGKGEVGV AEASKLSVEL IQQIRDDFEY WKRACQQYGA			AIGGGCTCCA	TAUTAGAWAG	TA ACCACCAT	САТСАСССАС	CTTGGTAAAT
TACATTGAAT CAGAGGTTGG AGCCAGGGTT GTGATACACA GCAAACCAGA CTATGACCT GAAAACAAGG CTGTGAATGC CGTTCCCCTG AAGCCAGCCA TATACCTTGA ATGA SEQ ID: 15 Archaeog lobus fulgidus leucyl trna-syntheta se (AFLRS) RS QWFLNYSNRE WKEKVLNHLE KMRIIPDYYK GAFLAG GAAACAAGG CTGTGAATGC CGTTCCCCTG AAGCCAGCCA AAGCCAGCCA CGTTCCCCTG AAGCCAGCCA CGTTCCCTG AAGCCAGCCA AAGCCAGCCA CGTTCCCCTG AAGCCAGCCA CGTTCCCCTG AAGCCAGCCA CGTTCCCCTG AAGCCAGCCA CGTTCCCCTG AAGCCAGCCA CGTTCCCCTG AAGCCAGCCA CGTTCCCCTG AAGCCAGCCA CGTTCCCCTG AAGCCAGCCA CGTTCCCCTG AAGCCAGCCA CGTTCCCCTG AAGCCAGCCA CGTTCCCCTG AAGCCAGCCA CGTTCCCCTG AAGCCAGCCA CGTTCCCCTG AAGCCAGCA CGTTCCCCTG AAGCCAGA CCAGCCA CGTTCCCCTG AAGCCAGCCA CGTTCCCCTG AAGCCAGCA CGTTCCCCTG AAGCCAGA CCAGACCAGA			GAAGGAGGTT	GCTGAATTTG	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	תאַרתראַתּכּתּאַ	ጥርሮልጥርጥርልጥ
CTATGACCCT GAAAACAAGG CTGTGAATGC CGTTCCCCTG AAGCCAGCCA TATACCTTGA ATGA SEQ ID: 15 Archaeog lobus fulgidus leucyl trna- syntheta se (AFLRS) RS QWFLNYSNRE KERVHEDLVK LGLGDVFYEF (AFLRS) RS QWFLNYSNRE LGTRIPWDKE LLGKGEVGKV AEASKLSVEL LGGKGEVGKV ALKSIGYSID WRRVFTTDE FTVIKFRLE DGDLIFPCAT LRPETVFGVT NIWVKPTTYV IAEVDGEKWF VSKEAYEKLT YTEKKVRLLE EVDASQFFGK YVIVPLVNRK VPILPAEFVD ALKSIGYSID WRRVFTTTDE HDQNPVEDHD LLAGEEATIV NIWVKPTTYV IAEVDGEKWF TDNATGVVMS VPAHAPFDLA AIEDLKRDEE TLAKYGIDKS VVESIKPIVL AIEDLKRDEE TLAKYGIDKS VVESIKPIVL GAFLRS) RS QWFLNYSNRE WKEKVLNHLE KMRIIPDYYK EEFRNKIEWL RDKACARRKG LGTRIPWDKE WLIESLSDST IYMAYYILAK YINAGLLKAE NMTPEFLDYV VLFHHVAIFP PDKWPRAIAV NGYVSLEGKK MSKSKGPLLT MKRAVQQYGA			CAGAGGTTAC	GGTGATAGAC	GAGTACAGCG	TUCTOTION	CCVVCVCV
SEQ ID: 15 Archaeog lobus fulgidus leucyl trna-syntheta se (AFLRS) RS QWFLNYSNRE KERVHEDLVK KERVHED	}		TACATTGAAT	CAGAGGTTGG	AGCCAGGGTT	GIGATACACA	A A CCC A CCC A
SEQ ID: 15 Archaeog lobus fulgidus leucyl trna-syntheta se (AFLRS) RS QWFLNYSNRE WCEKVLNHLE GTRIPWDKE (AFLRS) RS QWFLNYSNRE WCEKVLNHLE LLGKGEVGKV VLESKEVLH LLGKGEVGKV VLESKESVLL LLGKGEVGKV AEASKLSVEL LLGKGEVGKV VLEFHHVAIFP PDKWPRAIAV LGTRIPWDKE KGRAPRIT MRAKGYNVL FPLGFHVTGT PIIGLAELIA KRDERTIEVY ALKSIGYSID WRRVFTTTDE WRRVFTTTDE WRRVFTTTDE HDQNPVEDHD LLAGEEATIV LAGEEATIV LAGEEATIV VSKEAYEKLT YTEKKVRLLE EVDASQFFGK YVIVPLVNRK VPILPAEFVD AIEDLKRDEE TLAKYGIDKS VVESIKPIVL ATKTLYKKEY HTGIMLDNTM EKLIRELGVK SQKDKELLDK ATKTLYKKEY HTGIMLDNTM EFFRNKIEWL RDKACARRKG WKEKVLNHLE KMRIIPDYYK EEFRNKIEWL RDKACARRKG NMTPEFLDYV YLFHHVAIFP PDKWPRAIAV NGYVSLEGKK MSKSKGPLLT MKRAVQQYGA					CTGTGAATGC	CGTTCCCCTG	AAGCCAGCCA
1obus fulgidus leucyl trna- syntheta se (AFLRS) RS QWFLNYSNRE QWFLNYSNRE LGTRIPWDKE LLGKGEVGKV LLGKGEVGKV LLGKGEVGKV LLGKGEVGKV VYLFHHVAIFP PDKWPRAIAV VEYFSREALQ ALKSIGYSID WRRVFTTTDE VEYFSREALQ ALKSIGYSID WRRVFTTTDE VEYFSREALQ LLGLTPEKI VEYFSREALQ LLGLTPEKI VEYFSREALQ LLGLTPEKY NEYFOTT NIWVKPTTYV IAEVDGEKWF VVIVPLVNRK VPILPAEFVD ALGLGROUFYEF SEKPVICRCG TKCVVKVVRD RS VPLOTHOLONION RS TTIGDAFAR YMRMKGYNVL VEYFSREALQ LLGLTPEKI VEYFSREALQ ALKSIGYSID WRRVFTTTDE VEYFOTT NIWVKPTTYV IAEVDGEKWF VVIVPLVNRK VPILPAEFVD VVESIKPIVL ALGLGROUFYEF SEKPVICRCG TKCVVKVVRD RDKRAVQLEGKK MSKSKGPLLT MKRAVQQYGA			TATACCTTGA	ATGA			
1obus fulgidus leucyl trna- syntheta se (AFLRS) RS QWFLNYSNRE QWFLNYSNRE LGTRIPWDKE LLGKGEVGKV LLGKGEVGKV LLGKGEVGKV LLGKGEVGKV VYLFHHVAIFP PDKWPRAIAV VEYFSREALQ ALKSIGYSID WRRVFTTTDE VEYFSREALQ ALKSIGYSID WRRVFTTTDE VEYFSREALQ LLGLTPEKI VEYFSREALQ LLGLTPEKI VEYFSREALQ LLGLTPEKY NEYFOTT NIWVKPTTYV IAEVDGEKWF VVIVPLVNRK VPILPAEFVD ALGLGROUFYEF SEKPVICRCG TKCVVKVVRD RS VPLOTHOLONION RS TTIGDAFAR YMRMKGYNVL VEYFSREALQ LLGLTPEKI VEYFSREALQ ALKSIGYSID WRRVFTTTDE VEYFOTT NIWVKPTTYV IAEVDGEKWF VVIVPLVNRK VPILPAEFVD VVESIKPIVL ALGLGROUFYEF SEKPVICRCG TKCVVKVVRD RDKRAVQLEGKK MSKSKGPLLT MKRAVQQYGA	SEO ID: 15	Archaeog	MSDFRIIEEK	WQKAWEKDRI	FESDPNEKEK	FFLTIPYPYL	NGNLHAGHTR
fulgidus leucyl trna- syntheta se (AFLRS) RS QWFLNYSNRE CGTRIPWDKE LLGKGEVGKV LLGKGEVGKV LLGKGEVGKV LLGKGEVGKV LLGKGEVGKV VYLFHHVAIFP LLQLTTPEKI VEYFSREALQ ALKSIGYSID WRRVFTTTDE KGTHPVRYCP HDQNPVEDHD LLAGEEATIV LAGVEGETV NIWVKPTTYV IAEVDGEKWF VPILPAEFVD VPAHAPFDLA AIEDLKRDEE TLAKYGIDKS VVESIKPIVL AIEDLKRDEE TLAKYGIDKS VVESIKPIVL AIEDLKRDEE TLAKYGIDKS VVESIKPIVL AIEDLKRDEE TLAKYGIDKS VVESIKPIVL AIEDLKRDEE TLAKYGIDKS VVESIKPIVL AIEDLKRDEE TLAKYGIDKS VVESIKPIVL AIEDLKRDEE TLAKYGIDKS VVESIKPIVL AIEDLKRDEE TLAKYGIDKS VVESIKPIVL AIEDLKRDEE TLAKYGIDKS VVESIKPIVL AIEDLKRDEE TLAKYGIDKS VVESIKPIVL AIEDLKRDEE TLAKYGIDKS VVESIKPIVL AIEDLKRDE TLAKYGIDKS VVESIKPIVL AIEDLKRDEE TLAKYGIDKS VVESIKPIVL AIEDLKRDE TLAKYGIDKS AIEDLKRDE TLAKYGIDKS VVESIKPIVL AIEDLKRDE TLAKYGIDKS AIEDLKRDE TLAKYGIDKS AIEDLKRDE TLAKYGIDKS AIEDLKRDE TLAKYGIDKS AIEDLKRDE TLAKYGIDKS AIEDLKRDE T	<		TETTGDAFAR	YMRMKGYNVL	FPLGFHVTGT	PIIGLAELIA	KRDERTIEVY
EYQRFIEWQY WKLKELGLIV KGTHPVRYCP HDQNPVEDHD LLAGEEATIV trna- syntheta se (AFLRS) RS QWFLNYSNRE WKEKVLNHLE KMRIIPDYYK LGTRIPWDKE WLIESLSDST IYMAYYILAK LLGKGEVGKV YLFHHVAIFP PDKWPRAIAV NGYVSLEGKK MSKSKGPLLT MKRAVQQYGA			TKYHDVPLED	LLOLTTPEKI	VEYFSREALQ	ALKSIGYSID	WRRVFTTTDE
EFTVIKFRLE DGDLIFPCAT LRPETVFGVT NIWVKPTTYV IAEVDGEKWF VSKEAYEKLT YTEKKVRLLE EVDASQFFGK YVIVPLVNRK VPILPAEFVD Se IKTDIEGVPA EKLIRELGVK SQKDKELLDK ATKTLYKKEY HTGIMLDNTM NYAGMKVSEA KERVHEDLVK LGLGDVFYEF SEKPVICRCG TKCVVKVVRD RS QWFLNYSNRE WKEKVLNHLE KMRIIPDYYK EEFRNKIEWL RDKACARRKG LGTRIPWDKE WLIESLSDST IYMAYYILAK YINAGLLKAE NMTPEFLDYV LLGKGEVGKV AEASKLSVEL IQQIRDDFEY WYPVDLRSSG KDLVANHLLF YLFHHVAIFP PDKWPRAIAV NGYVSLEGKK MSKSKGPLLT MKRAVQQYGA		_	EVORETEMOV	WKIKELGITV	KGTHPVRYCP	HDQNPVEDHD	LLAGEEATIV
SYNTHETA SYNTHETA SYNTHETA SE (AFLRS) RS (AFRCS) RS (AF		-	PIONTIEWOI P	DCDI.TEDCAM	TRPENVECON	NIWVKPTTYV	IAEVDGEKWF
SYNTHETA SE (AFLRS) RS QWFLNYSNRE WKEKVLNHLE KMRIIPDYYK EEFRNKIEWL RDKACARRKG LGTRIPWDKE WLIESLSDST IYMAYYILAK YINAGLLKAE NMTPEFLDYV LLGKGEVGKV AEASKLSVEL IQQIRDDFEY WYPVDLRSSG KDLVANHLLF YLFHHVAIFP PDKWPRAIAV NGYVSLEGKK MSKSKGPLLT MKRAVQQYGA		trna-	FLIATVLKTE	DODDILL LCW1	TITI DI ALGAI	VIJTI/DI.IMIPK	VPILPAERVD
SE IKTDIEGVPA EKLIRELGVK SQKDKELLDK ATKTLYKKEY HTGIMLDNTM (AFLRS) RS QWFLNYSNRE WKEKVLNHLE KMRIIPDYYK EEFRNKIEWL RDKACARRKG LGTRIPWDKE WLIESLSDST IYMAYYILAK YINAGLLKAE NMTPEFLDYV LLGKGEVGKV AEASKLSVEL IQQIRDDFEY WYPVDLRSSG KDLVANHLLF YLFHHVAIFP PDKWPRAIAV NGYVSLEGKK MSKSKGPLLT MKRAVQQYGA		1	VSKEAYEKLT	X.LEKKAKPPE	PATENT KEDEL	TATA ETIANNY TATA ETIANNY	WECTKOTIT.
(AFLRS) RS QWFLNYSNRE WKEKVLNHLE KMRIIPDYYK EEFRNKIEWL RDKACARRKG LGTRIPWDKE WLIESLSDST IYMAYYILAK YINAGLLKAE NMTPEFLDYV LLGKGEVGKV AEASKLSVEL IQQIRDDFEY WYPVDLRSSG KDLVANHLLF YLFHHVAIFP PDKWPRAIAV NGYVSLEGKK MSKSKGPLLT MKRAVQQYGA		_	TDNATGVVMS	VPAHAPFDLA	ALEDUKRDEE	THAKIGIDKS	A A EDITULITATION
RS QWFLNYSNRE WKEKVLNHLE KMRIIPDYYK EEFRNKIEWL RDKACARRKG LGTRIPWDKE WLIESLSDST IYMAYYILAK YINAGLLKAE NMTPEFLDYV LLGKGEVGKV AEASKLSVEL IQQIRDDFEY WYPVDLRSSG KDLVANHLLF YLFHHVAIFP PDKWPRAIAV NGYVSLEGKK MSKSKGPLLT MKRAVQQYGA			IKTDIEGVPA	EKLIRELGVK	SQKDKELLDK	ATKTLYKKEY	HIGIMUDNIM
RS QWFLNYSNRE WKEKVLNHLE KMRIIPDYYK EEFRNKIEWL RDKACARRKG LGTRIPWDKE WLIESLSDST IYMAYYILAK YINAGLLKAE NMTPEFLDYV LLGKGEVGKV AEASKLSVEL IQQIRDDFEY WYPVDLRSSG KDLVANHLLF YLFHHVAIFP PDKWPRAIAV NGYVSLEGKK MSKSKGPLLT MKRAVQQYGA			NYAGMKVSEA	KERVHEDLVK	LGLGDVFYEF	SEKPVICRCG	TKCVVKVVRD
LGTRIPWDKE WLIESLSDST IYMAYYILAK YINAGLLKAE NMTPEFLDYV LLGKGEVGKV AEASKLSVEL IQQIRDDFEY WYPVDLRSSG KDLVANHLLF YLFHHVAIFP PDKWPRAIAV NGYVSLEGKK MSKSKGPLLT MKRAVQQYGA		RS	OWELNYSNRE	WKEKVLNHLE	KMRIIPDYYK	EEFRNKIEWL	RDKACARRKG
LLGKGEVGKV AEASKLSVEL IQQIRDDFEY WYPVDLRSSG KDLVANHLLF YLFHHVAIFP PDKWPRAIAV NGYVSLEGKK MSKSKGPLLT MKRAVQQYGA	<u>[</u>		LGTRIPWDKE	WLIESLSDST	IYMAYYILAK	YINAGLLKAE	NMTPEFLDYV
YLFHHVAIFP PDKWPRAIAV NGYVSLEGKK MSKSKGPLLT MKRAVQQYGA	[I'I'CKCEACKA	AEASKI SVET	IOOIRDDFEY	WYPVDLRSSG	KDLVANHLLF
DVTRLYILHA AEYDSDADWK SREVEGLANH LRRFYNLVKE NYLKEVGELT	1		AL ELITAY TED	וזע ב ע מפואאטם	MCANCI ECKK	MSKSKGPLLT	MKRAVOOYGA
DVTRLYILHA AEYUSDADWA SKEVEGLANA DAAF MIDAEVGEDI]		Y LFHHVAIFP	FUMBRATHA	GDEALCA VILL MOTA PIERVI	T'BBEAML'ME	MAI'KEAGEI'A
	1		DVTRLYILHA	AEYUSUADWK	OKEVEGUANH	DVVI. LIATIA VE	WITHURDAGE

γ			OD A THE TOPA	MONT OFFI	MAN DEPET MAID	VIDIATYT DDCCE
I		TLDRWLVSRM	QRAIKEVREA	MDNLQTRKAV	NAAFFELMND	AUMITIVIGGE
		MINITIPOPMI	KLLAPFAPHI	CEELWHLKHD	SYVSLESYPE	IDEIRVDEEA
		ERIEEYLRNL	VEDIQEIKKF	VSDAKEVYIA	PAEDWKVKAA	KVVAESGDVG
					KKLMLVKEWE	APŐÖNPKETE
		NETGLKVILD	TORVPEEKRR	QAVPGKPAIY	VA*	
SEQ ID: 16	Methanob	VDIERKWRDR	WRDAGIFQAD	PDDREKIFLT	VAYPYPSGAM	HIGHGRTYTV
	acterium	PDVYARFKRM	QGYNVLFPMA	WHVTGAPVIG	IARRIQRKDP	WTLKIYREVH
	thermoau	RVPEDELERF	SDPEYIVEYF	SREYRSVMED	MGYSIDWRRE	FKTTDPTYSR
-					PVGDHDLLEG	
	totrophi	LKEKLGDSYL	VAATERPETI	YGATNLWLNP	DEDYVRVETG	GEEWIISRAA
1	cum	VIDNITISHOKIJD	T.KVSGDVNPG	DITEMOVENE	VTGQEHPILP	ASFVDPEYAT
	leucyl	CIMEGINDAHA	DANFTALENI.	RADHELLERA	GLEDVVADIE	PVNVIAVDGY
	trna-	GAALDALWILL	KECKDNOEDD	PLEDATCELV	KIEHARGVMS	SHTPVYGGMK
!	syntheta	GELLAVEATE	KEGAKNOEDE	MARCHEDOLLI	CRCGGRCVVR	MEDOMEMKY
	se	ASEAKEATAD	EDVDÖGDYDE	MIELWELLAT	IDWLNDWACS	סדבו.כייפו.ם
		SDDAWKDLAH	RCLDGMKIIP	EEVKAMPEII	TOWNDAMACS	VYTGDGTIGH
·	(MtLRS)	WDERWIIEPL	TDSTIYMAYY	TIAHRLREMD	AGEMDDEFFD	AIFLUDSGIF
		EDLREEFRYW	YPLDWRLSAK	DLIGNHLTFH	IFHHSAIFPE	SGWPRGAVVF
		GMGLLEGNKM	SSSKGNVILL	-RDAIEKHGAD	VVRLFLMSSA	EPWQDFDWRE
		SEVIGTRRRI	EWFREFGERV	SGILDGRPVL	SEVTPAEPES	FIGRWMMGQL
l		NQRIREATRA	LESFQTRKAV	QEALYLLKKD	VDHYLKRVEG	RVDDEVKSVL
		ANVLHAWIRL	MAPFIPYTAE	EMWERYGGEG	FVAEAPWPDF	SDDAESRDVQ
		VAEEMVQNTV	RDIQEIMKIL	GSTPERVHIY	TSPKWKWDVL	RVAAEVGKLD
		MGSIMGRVSA	EGIHDNMKEV	AEFVRRIIRD	LGKSEVTVID	EYSVLMDASD
			VIHSKPDYDP			
SEQ ID: 17	DACKO-	gaactccgga	tgagcattca	tcaggcgggc	aagaatgtga	ataaaggccg
SEQ ID. 17	-	gaaccccgga	atacttattt	ttctttacgg	tctttaaaaa	ggccgtaata
	A184AGGA	tagagataaa	caatctaatt	ataggtagat	tgagcaactg	actgaaatgc
		ctcagctgaa	tatttagget	acaytacae	tatatcaacg	gtggtatatc
		Cicadaatgi	tetetacyat	ttagettagt	tagctcctga	asatctccat
		cagtgatttt	ttteteeatt	tratertatt	atttcattat	aataaaatt
		aactcaaaaa	atacgcccgg	tagigatett	ttttagggaa	agetteece
		ggaacctctt	acgraccat	caacgtetea	ttttcgccaa	ttatagassa
	,	agggcttccc	ggtatcaaca	gggacaccag	gatttattta	caratastas
		tgatcttccg	tcacaggtat	ttatteggeg	caaagtgcgt	cygycyacyc
		tgccaactta	ctgatttagt	gtatgatggt	gtttttgagg	tgeteeagtg
		gcttctgttt	ctatcagctg	tccctcctgt	tcagctactg	acggggtggt
		gcgtaacggc	aaaagcaccg	ccggacatca	gcgctagcgg	agtgtatact
		ggcttactat	gttggcactg	atgagggtgt	cagtgaagtg	cttcatgtgg
		caggagaaaa	aaggctgcac	cggtgcgtca	gcagaatatg	tgatacagga
1		tatattccgc	ttcctcgctc	actgactcgc	tacgctcggt	cgttcgactg
		cggcgagcgg	aaatggctta	cgaacggggc	ggagatttcc	tggaagatgc
1		caggaagata	cttaacaggg	aagtgagagg	gccgcggcaa	agccgttttt
		ccataggete	egecececta	acaagcatca	cgaaatctga	cgctcaaatc
	·	agtggtggcg	aaacccgaca	ggactataaa	gataccaggc	gtttccccct
		adcaactccc	teatacacte	tcctattcct	gcctttcggt	ttaccggtgt
		catteegete	tratogogo	gtttgtctca	ttccacgcct	gacactcagt
		teceastes	canttenete	caarctrosc	tgtatgcacg	aaccccccat
1		tasatasas	cagticigoto	tateceetaa	ctatcgtctt	gagtccaacc
1		ccagcccgac	tacocca	agactarasa	carcacter	taattoattt
Į.		cggaaagaca	rgcaaaagca	ccactggcag	cagccactgg	actosasons
		agaggagtta	gtcttgaagt	catgegeegg	ttaaggctaa	attassaca
1		caagttttgg	tgactgcgct	cctccaagcc	agttacctcg	guudaayay
Ī		ttggtagctc	agagaacctt	cgaaaaaccg	ccctgcaagg	cggtttttc
]		gttttcagag	caagagatta	cgcgcagacc	aaaacgatct	caagaagatc
1	1	atcttattaa	tcagataaaa	tatttctaga	tttcagtgca	atttatctct
!	ĺ	tcaaatgtag	cacctgaagt	cagccccata	cgatataagt	tgtaattctc
		atgtttgaca	gcttatcatc	gataagcttt	aatgcggtag	tttatcacag
1		ttaaattgct	aacgcagtca	ggcaccgtgt	atgaaatcta	acaatgcgct
1		catcotcatc	ctcaacacca	tcaccctgga	tgctgtaggc	ataggcttgg
		ttatgccggt	actoccoooc	ctcttacaaa	atatcGGTTT	CTTAGACGTC
		AGGTCCCACT	TTTCCCCSSSC	atatacacaa	aacccctatt	tgtttatttt
L	L	LAGGIGGCACI	11 ccggggaa	2030303039		

tctaaataca ttcaaatatg tatccgctca tgagacaata accctgataa atgcttcaat aatattgaaa aaggaagagt atgagtattc aacatttccg tqtcgccctt attccctttt ttgcggcatt ttgccttcct gtttttgctc acccagaaac gctggtgaaa gtaaaagatg ctgaagatca gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga tccttgagag ttttcgcccc gaagaacgtt ttccaatgat gagcactttt aaagttctgc tatgtggcgc ggtattatcc cgtgttgacg ccgggcaaga gcaactcggt cgccgcatac actattctca gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg gcatgacagt aagagaatta tgcagtgctg ccataaccat gagtgataac actgcggcca acttacttct gacaacgatc ggaggaccga aggagctaac cgcttttttg cacaacatgg gggatcatgt aactcgcctt gatcgttggg aaccggagct gaatgaagcc ataccaaacg acqaqcqtqa caccacqatg cctAGGAgca atggcaacaa cgttgcgcaa actattaact ggcgaactac ttactctagc ttcccggcaa caattaatag actggatgga ggcggataaa gttgcaggac cacttctgcg ctcggccctt ccggctggct ggtttattgc tgataaatct ggagccggtg agcgtgggtc tcqcqqtatc attqcaqcac tggggccaga tggtaagccc tcccgtatcg tagttatcta cacgacgggg agtcaggcaa ctatggatga acgaaataga cagatcgctg agataggtgc ctcactgatt aagcattggc accaccacca ccaccactaa CCCGGGACCA AGTTTACTCA TATATACttt agattgattt aaaacttcat ttttaattta aaaggatcta ggtgaagatc ctttttgata atCTCATGAC CAAAATCCCT TAACGgcatg caccattcct tgcggcggcg gtgctcaacg gcctcaacct actactGGGC TGCTTCCTAA TGCAGGAGTC GCATAAGGGA GAGCGTCTGG CGAAAGGGGG ATGTGCTGCA AGGCGATTAA GTTGGGTAAC GCCAGGGTTT TCCCAGTCAC GACGTTGTAA AACGACGGCC AGTGCCAAGC TTAAAAAaaa tccttagctt tcgctaagga tCTGCAGTTA TAATCTCTTT CTAATTGGCT CTAAAATCTT TATAAGTTCT TCAGCTACAG CATTTTTTAA ATCCATTGGA TGCAATTCCT TATTTTTAAA TAAACTCTCT AACTCCTCAT AGCTATTAAC TGTCAAATCT CCACCAAATT TTTCTGGCCT TTTTATGGTT AAAGGATATT CAAGGAAGTA TTTAGCTATC TCCATTATTG GATTTCCTTC AACAACTCCA GCTGGGCAGT ATGCTTTCTT TATCTTAGCC CTAATCTCTT CTGGAGAGTC ATCAACAGCT ATAAAATTCC CTTTTGAAGA ACTCATCTTT CCTTCTCCAT CCAAACCCGT TAAGACAGGG TTGTGAATAC AAACAACCTT TTTTGGTAAA AGCTCCCTTG CTAACATGTG TATTTTTCTC TGCTCCATCC CTCCAACTGC AACATCAACG CCTAAATAAT GAATATCATT AACCTGCATT ATTGGATAGA TAACTTCAGC AACCTTTGGA TTTTCATCCT CTCTTGCTAT AAGTTCCATA CTCCTTCTTG CTCTTTTTAA GGTAGTTTTT AAAGCCAATC TATAGACATT CAGTGTATAA TCCTTATCAA GCTGGAATTC agcgttacaa gtattacaca aagtttttta tgttgagaat attttttga tggggcgcca cttatttttg atcgttcgct caaagAAGCG GCGCCAGGGN TGTTTTCTT TTCACCAGTN AGACGGGCAA CAGAACGCCA TGAgcggcct cattlcttat tctgagttac aacagtccgc accgctgtcc ggtagctcct tccggtgggc gcggggcatg actatcgtcg ccgcacttat gactgtcttc tttatcatgc aactcgtagg acaggtgccg gcagcgccca acagtccccc ggccacgggg cctgccacca tacccacgcc gaaacaagcg ccctgcacca ttatgttccg gatctgcatc gcaggatgct gctggctacc ctgtggaaca cctacatctg tattaacgaa gcgctaaccg tttttatcag gctctgggag gcagaataaa tgatcatatc gtcaattatt acctccacgg ggagagcctg agcaaactgg cctcaggcat ttgagaagca cacggtcaca ctgcttccgg tagtcaataa accggtaaac cagcaataga cataagcggc tatttaacga ccctgccctg aaccgacgac cgggtcgaat ttgctttcga atttctgcca ttcatccgct tattatcact tattcaggcg tagcaccagg cgtttaaggg caccaataac tgccttaaaa aaattacgcc ccgccctgcc actcatcgca gtactgttgt aattcattaa gcattctgcc gacatggaag ccatcacaga cggcatgatg aacctgaatc gccagcggca tcagcacctt gtcgccttgc gtataatatt tgcccatggt gaaaacgggg gcgaagaagt tgtccatatt ggccacgttt aaatcaaaac tggtgaaact cacccaggga ttggctgaga cgaaaaacat attctcaata aaccctttag ggaaataggc caggttttca ccgtaacacg ccacatcttg cgaatatatg tgtagaaact gccggaaatc

WO 2005/007870

						· · · · · · · · · · · · · · · · · · ·
		gtcgtggtat	tcactccaga	gcgatgaaaa	cgtttcagtt	tgctcatgga
		aaacggtgta	acaagggtga	acactatccc	atatcaccag	ctcaccgtct
		ttcattgcca				
		_				
SEQ ID: 18	DACKO-	gaactccgga	tgagcattca	tcaggcgggc	aagaatgtga	ataaaggccg
SEQ ID. 10		gatecoogga	gtgcttattt	ttctttacgg	tctttaaaaa	ggccgtaata
	A184TAG	tacaactaaa	castctaatt	atacctacat	tgagcaactg	actgaaatgc
		-tananatat	tatttacast	acaggiacas	tatatcaacg	gtggtatatc
		cccaaaacgc	thetacyat	ttacattcct	tagctcctga	aaatctcgat
-		cagtgattt	ttttttttttt	teagetteet	atttcattat	adtasaaatt
		aactcaaaaa	atacgcccgg	caytyatte	atttcattat	aarttraccc
	1	ggaacctctt	acgtgccgat	caacytetta	ttttcgccaa	ttctcccaac
		agggcttccc	ggtatcaaca	gggacaccay	gatttattta	ttctgcgaag
		tgatcttccg	tcacaggtat	ttattcggcg	caaagtgcgt	testessets
		tgccaactta	ctgatttagt	gtatgatggt	gtttttgagg	tgctccagtg
		gcttctgttt	ctatcagctg	tccctcctgt	tcagctactg	acggggtggt
	i	gcgtaacggc	aaaagcaccg	ccggacatca	gcgctagcgg	agtgtatact
		ggcttactat	gttggcactg	atgagggtgt	cagtgaagtg	cttcatgtgg
		caggagaaaa	aaggctgcac	cggtgcgtca	gcagaatatg	tgatacagga
		tatattccgc	ttcctcgctc	actgactcgc	tacgctcggt	cgttcgactg
		cggcgagcgg	aaatggctta	cgaacggggc	ggagatttcc	tggaagatgc
!		caggaagata	cttaacaggg	aagtgagagg	gccgcggcaa	agccgttttt
		ccataggctc	cacccccta	acaagcatca	cgaaatctga	cgctcaaatc
		agtggtggcg	aaacccgaca	ggactataaa	gataccaggc	gtttccccct
	1	ageggeteee	tcatacacte	tcctattcct	gcctttcggt	ttaccggtgt
		cattccctc	ttataaccac	ghttgtctca	ttccacacact	gacactcagt
	}	taccegecg	cartterete	caagetggae	totatocaco	aaccccccgt
		tecgggtagg	cagteegeet	tatcccctaa	ctatcgtctt	gagtccaacc
		ceagiceegae	teassages	gazatagaa	carccactaa	taattgattt
		cggaaagaca	Lycadaayca	ccaccggcag	ttaaggctaa	actgaaagga
		agaggagtta	gccccgaage	catgegeegg	agttaggetag	actgaaagga
		caagttttgg	tgactgcgct	cetecaagee	agetaceteg	gttcaaagag
		ttggtagctc	agagaacctt	cgaaaaaccg	ceeegeaagg	cggttttttc
		gttttcagag	caagagatta	cgcgcagacc	aaaacgaccc	caagaagatc
İ	-	atcttattaa	tcagataaaa	tatttctaga	tttcagtgca	atttatctct
		tcaaatgtag	cacctgaagt	cagccccata	cgatataagt	tgtaattctc
		atgtttgaca	gcttatcatc	gataagcttt	aatgcggtag	tttatcacag
		ttaaattgct	aacgcagtca	ggcaccgtgt	atgaaatcta	acaatgcgct
	1	catcgtcatc	ctcggcaccg	tcaccctgga	tgctgtaggc	ataggcttgg
		ttatgccggt	actgccgggc	ctcttgcggg	atatcGGTTT	CTTAGACGTC
		AGGTGGCACT	TTtcggggaa	atgtgcgcgg	aacccctatt	tgtttatttt
	1	tctaaataca	ttcaaatatg	tatccgctca	tgagacaata	accctgataa
	1	atgcttcaat	aatattgaaa	aaggaagagt	atgagtattc	aacatttccg
		tateaccett	attccctttt	ttgcggcatt	ttgccttcct	gtttttgctc
		acccagaaac	gctggtgaaa	gtaaaagatg	ctgaagatca	gttgggtgca
		cgagtgggtt	acatcgaact	ggatctcaac	agcggtaaga	tccttgagag
	1	ttttcacccc	gaagaacgtt	ttccaatgat	gagcactttt	aaagttctgc
		tatatagacac	ggtattatcc	catattaaca	ccgggcaaga	gcaactcggt
1		cacacacatac	actattctca	gaatgacttg	attaaatact	caccagtcac
		agaaaaggat	cttaccccca	gcatgacagt	aagagaatta	tgcagtgctg
1		ayaaaaycat	rantratare	actorogo	acttacttct	gacaacgatc
		CCataaccat	yayıyatadü	cacttttt	cacaacatoo	gggatcatgt
		ggaggaccga	aggagetaae	22222222	rastrascr	ataccaaaco
!		aactcgcctt	gaccgccggg	aaccyyayct	tarceaces	ataccaaacg
1		acgagcgtga	caccacgatg	cctracgcaa	tageagass	gttgcgcaaa
		ctattaactg	gcgaactact	tactctagct	cccggcaac	aattaataga
1		ctggatggag	gcggataaag	ttgcaggacc	acttetgege	teggeeette
		cggctggctg	gtttattgct	gataaatctg	gagccggtga	gcgtgggtct
		cocootatca	ttgcagcact	ggggccagat	ggtaagccct	cccgtatcgt
1		agttatctac	acqacqqqqa	gtcaggcaac	tatggatgaa	cgaaatagac
		agategetga	gataggtgcc	tcactgatta	agcattggca	ccaccaccac
		caccactaaC	CCGGGACCAA	GTTTACTCAT	ATATACttta	gattgattta
L	<u> </u>	· · · · · · · · · · · · · · · · · · ·				

Basctteat titaattaa aggatctag glagagatcs tittigataa ticttotageg cotcaacota clacicagot gotcaacota clacicagot gotcaacota clacicagota Gotcotacota control gotcotacota control gotcotacota control gotcotacota control gotcotacota control gotcotacota control gotcotacota control gotcotacota control gotcotacota control gotcotacota control gotcotacota control gotcotacota control gotcotacota control gotcotacota control gotcotacota control gotcotacota control gotcotacotacota control gotcotacotacotacotacotacotacotacotacotaco			agacttcatt	tttaatttaa	aaggatctag	gtgaagatcc	tttttgataa
Egiclasagg cotoaaccta chactGGGT GTTCCTATA GGAGGAGTG CATAAGGAGA AGCGTTTAC CCAAGTCAA GGCATTAAC TTGGTAACG CAAGGGTTT CCAATCACC AGTGTTAAA AGCAGGCCA TAAAAAaaa cottagett cgcaaggag tCGCAGTTAA AATCTTTTAA TCATTGGTC TAAAATCTT ATAAGTTCT CACTACAGC ATTTTTAAA TCATTGGAT GCAATCTT ATAAGTTCT CACTACAGC TTATTGGTA AAGAATTAC ACAACTCCAA GCCATATTGAA ATTTTTTAAA TCATTGGAT GCAATCTT ATAAGTTCT CACTACAGC ATTTCCTTCT AAGAATTAT CAAAACTCT AACAACTCTAA ATTTCTTCT TGGAGAGTA TCAAACGCT TTTTAAGAAATT TTTGCGGCTT TAATCTCTC TGGAGAGTCA TCAACACCTA TAAAATTCC CACAAATT TTTGCGGCT TAACCACCT TTTGGATAAA GCTCCCTGC AACAACCTT TTTGGATAAAA GCTCCCTTGC AACAACCTT TTTGGATAAAA GCTCCCTTGC AACAACCTT TTTGGATAAAA ACTCCAGC AACACCGT TTTGAAGAA ACAACCTT TTTGGATAAAA ACTCCAGC AACACCGT TTTTAAGACA ACAACCTAC TTGGATAGAA ACTCAAGC CAAACCGT TTTTAAGACA ACACCTGATTA TGGATAGAA TACTCAGC ACCTTGGAT TTTCATCCTC TCTTGCTATA AGTCCATAC ACATCAAC ACCTTAGAT TTTCATCTC TCTTGCTATA AGTCCATAC ACGTCATACTAGA ACCTTTGAAT TATTCTTCT TCTTGCTATA AGTCCATAC ACGTCATACCA ACCTTAGAT TTTCATCCTC TCTTGCTATA AGTCCATAC ACGTCATACCA ACCTTAGAT TTTCATCCTC TCTTGCTATA AGTCCATAC ACGTCATACCA ACCTTTGGAT TTTCATCCTC TCTTGCTATA AGTCCATAC ACGTCATACCA ACCTTAGAT TTTCATCCTC TCTTGCTATA AGTCCATAC ACGTCATACCA ACCTTAGAT TTTCATCCTC TCTTGCTATA AGTCCATAC ACGTCATACCA ACCTTTGGAT TTTCATCCTC TCTTGCTATA AGTCCATAC ACGTCATACCA ACCTTTGGAT TTTCATCCTC TCTTGCTATA AGTCCATAC ACCTTTTCTTTC CCAATGAAACACCA AGAACCCAT GGGGGATCA TCTTTTCTTT TCACCAATAA ACTCCAACCCA AGAACCCAT GGGGGGCAC TCTTTTCTTT TCACCAATAA ACCTCATACCA ACCTTTAGAA TACGAACTCA ACGTTATATGAAATTACACATTC ACCTTTTAAAAAAAAAA			+CTCATGACC	AAAATCCCTT	AACGgcatgc	accattcctt	acaacaacaa
CATARAGGAN ACCTOTAGE GAAAGGGGA TOTGETGAA ACGCATTAMA TTGGTAAC CCAGGGTTT CCCAGTCACA ACGTTGTAA ACGCAGGCA GTGCCAAGCT TAAAAATCTT CCAGTCACA ACGTTGTAAA ACGCAGGCA ATTTTTTAAA TCATTGGTC TAAAATCTT ATAAGTTCT CACTACACC ATTTTTTAAA TCCATTGAT GCAATTCTT ATATGTTCT CACTACACC ATTTTTTAAA TCCATTGAT GCAATTCTT ATATGTTCT CACTACACC ATTTATCTTCA ACGACTCCAG CTGGGCATA TACTTTTTA TTTGGCCTT TTATGGTTA AAGGATTCT CAACACCCTA TAAATTCC TTCTAGCCT TAATTCTCTC TGGAGAGTA TAAACTCCTA CAAACCCGT AAGACAGGT TATAAGGATACC TCATCTTCC CTCTCCATC CAAACCCGTT AAGACAGGT TGGAATACC ACCTCATCCT CCAACTCCA ACATCAAACCC TAAAATACC TTTTAGAGA ACCACCTT TTTGGTAAAA GCTCCTTGC TAAACACTGT ATATTCCTT ACCACCATC TCCACTCCA ACATCAACCC CTAAATATG ATTTCCTTC TCTTGCTATA AGTTCCATAC TCCTTCTCC TTTTTTAAG GTAGTTTTTA AACCCAATCT ATTGGATAGAT AACTTCACCA ACCTTGTAA TTTTCTCTT AAGCCAATCT ATTGGATAGAT AACTTCACCA ACCTTTATAG GTAGTTTTTA AAGCCAATCT ATTGGATAGAT CACTTCACC CTTAATATG GTAGTTTTTA AAGCCAATCT ATTGACATTC AGTGTATAAT CCTTTTTAAG GTAGTTTTTA AAGCCAATCT ATTGACATTC AGTGTATAAT CCTTTTTAAG GTAGTTTTTA AAGCCAATCT ATTGACATTC AGTGTATAAT CCTTTTTAAG GTAGTTTTTA AAGCCAATCT ATTGACCATT AAGTGTATAAT CCTTTTTAAG GTAGTTTTTA AAGCCAATCT ATTGACCATT AAGTGTAATAT CCTTTATAGG GTAGTTTTTA AAGCCAATCT ATTGACCATT AAGTGTATAT CCTTTTTAAG GTAGTTTTTA AAGCCAATCT ATTGACCATT AAGTGTATAT CCTTTTTAAG GTAGTTTTTA AAGCCAATCT ATTGACCATT AAGTGTATAT CCTTTTTAAG GTAGTTTTTA AAGCCAATCT ATTGACCATT AAGTGTATAT CCTTTTTAAG GTAGTTTTTA AAGCCAATCT ATTGACATTC AGTGTATATAT CCTTTTTAAG GTAGTTTTTA AAGCCAATCT TACACAATTA AGCGGAACA GAAACCGCT GAGGGACTA GTTTTTCTTT TCACCAATTA AACCACATC GAGGTATATAT CCTTTTTAAG GTGGTATTTCTTT TCACCAATTA AACCACACACACACAACAACACCCACC			tactuature	cctcaaccta	ctactGGGCT	GCTTCCTAAT	GCAGGAGTCG
TTGGGTAACG CCAGGGTTT GTCACAGCACCA GTGCCAACCT TAAATGCTTC TAATTGGCTC TAAATTCTTTC TAATTGGCTC TAAATTCTTT TATTTTAAAT TCACTCCATA ACTCCTCATA GCATTACTT TATTAGGTTA ACCCCTCATA GCATTACT ATTTTTTAAAT ACCCCTCATA ACTTCTTCA ACCCCCCATA ACTTCTTCA ACCCCCCATA ACCCCCCATA ACCCCCCATA ACCCCCCATA ACCCCCCATA ACCCCCCATA ACCCCCCATA ACCCCCCATA ACCCCCCATA ACCCCCCATA ACCCCCCATA ACCCCCCATA ACCCCCCATA ACCCCCCATA ACCCCCCCATA ACCCCCCATA ACCCCCCATA ACCCCCCATA ACCCCCCCATA ACCCCCCATA ACCCCCCCATA ACCCCCCCTC TAATCCCTC TGCACCCCA ACCCCCACCC ACACCCCATA ACCCCCCCTC ACCCCCACCC ACCCCCACCC ACCCCCACCC ACCCCCACCC ACCCCCC	1	,	CAMAACCCAC	ACCCTCTCCC	CANACCCCCA	TCTCCTCCAA	GGCGATTAAG
GTGCCAGCT TANANAGAR ANTOTOTITO TANITHGOTO TANANTOTTO CACTACAGO ATTITITANA TCCATTGAT GCAATTCT ATATGATTCT CACTACAGC ATTITITANA TCCATTGAT GCAATTCT ATATTTANAT AACTCTTA ACTCCTCATA GTATATACT GTAAATTCT CACCAAATTT TTTATGGTA ATTICCTTCA ACAACTCCAG CTGGCAGTA TCCATTATT TATAGTATCC TAATTCCTTC TGGAGAGTAT CACAAGGTA TATATCCC TTTAGGCC TAATTCCTTC CTGCAGCACTA CACAAGGTA TATATCCC TTTAGAGA ACAACCTT TTTGGTAAAA GCTCCCTTG TAAAATACC TTTAGAGA ACAACCATT TTTGGTAAAA GCTCCCTTG TAAAATACC TTTGAGAGA ACAACCATT TTTGGTAAAA GCTCCCTTG TAAAATACT TTTCTCTTCAT GCTCCATCC TCCAACTGCA ACATCAGCA ACCTTTGATATTTCTCTT ACCTGCATTA TAGGTAGAT AACTCCAGC TAAAATACT ATTCATCCTC TCTTGCATA AGTTCCATAC ACATCAGCA ACCTTTGAGA ACTTCATCA AACCCAATCT ATAGGACATC AGTGTATAAT CCTTTTTAGG GTAGTTTTTA ACCTGCATTA TAGGTAGAT ACTTCAGCA ACCTTTGAGA ACTGCATA AAGCCAATCT ATAGGACATC AGTGTATAAT CCTTTTTAGG GTAGTTTTTA ACCTGCATTA TAGGTAGAT ACTCTCTTGC TCTTGCATA AGTTCCATAC ACATCAGCA ACCAACGCAT GAGGTTTTTC ACAGCATTT TACACCATTA AGTTCCATAC GGTTTTTCTT TCACCATATA ACCTCAGCA ACAACGCCCT GAGGTTTTCACACTT TCACCACATATA ACCTGCAAA ACAACGCCTA GAGGTTTTACACCTC TCTTGCTATA AGTTCCATAC ACTTCAGCA ACCATTGAGA TCCATCCTC TCTTGCATAT AGTTCCATAC ACTTCAGCA ACAACGCCCT GAGGTTTTCACACTT TCACCACATATA ACCTGTATAAA CCTGAGCATA GAAACCGCT GAGGTTTTCACACTATT AAACCCATACACACACACAACAAAACCGCTA GAGGTTTTCACACATATA ACCTGACACAACAAAACCGCCTA GAGGTATTCAACACACACACAACAAAACCACCATACAACAACAA							
AATCCTTTC TAATTGCCT TAAATCTT ATTAATAA AACTCTCAA ACTCCTCATA GCTATTAACT GCAAATCT ATTTTAAAA AACTCTCTAA ACTCCTCATA GCTATTAACT GTCAAATCT CACCAAATTT TTCTGGCCTT TTTATGGTTA AAGAGAATTT CAGCAGAGTATT TATGCTATCT CATTATTGG ATTTCCTCA ACAACTCCAG CTGGCGACTA TCCTTTCTTT ATCTTAGCCC TAAATCTCTC TGTGAGAATCA CAAACAGCTA TAAAATTCCC TTTTGAAGAA CCTCATCTT CTTCTCAAC CAAACCGCT AAGAAGGGT TGTGAATACA AACAACCTT TTTGGTAAAA GCTCCTTGC TAACATGTG ATTTTCCTTC ACCTCCATTA TTGGATAAA GCTCCTTGC TAACATGTG ATTTTCCTTC ACCTCCATTA TTGGATAAA GCTCTTCAGC CTAAATACA ATTTTCCTTC TCTTCCTATA ACTTCCATAC TCCTTCTTCT CTTTTTTAAG GAATCA GCGAATCA TAGGACATC ACATCACA ACCTTCAAC TCCAATCCAG GCGTCATCAT ATGGACTACA TCCTTCTTCT CTTTTTTAAG GAAGTCATC GCGTTCATCT ATGGACTAC TCCTTCTTCT CTTTTTTAAG GAAGTCATCA GCGTTTTCTTT TCACCAGTAA GAGGGCAAC CAAAACGCCAT GTTTTCTTT TCACCAGTAA GAGGGCAAC CAAAACGCCAT GAGGGGCAC GTTTTTCTTT TCACCAGTAA GAGGGCAAC CAAAACGCCAT GAGGGCAC GCCAGGGGT GGGCCCCC CAGGGGGGGGGCCCCCACCGGGTT GAGGACGCAC GAGACGCCAT GAGGGCAC CACCACCGGGT GGGCCCCCGGGT GGGCCCCCGGGT GGGCCCCCGGGT GGGCCCCCGGGT GGGCCCCCGGGT GGGCCCCCGGGT GGCCCCGGGT GGCCCCGGGT GGCCCCGGGT GGCCCCGGGGT GGCCCCGGGCCCCGGGCCCGGCCCGCGCCGCCCGC							
ACTOTTOTA GOTATIAC GCAATTCCTT ATTITAAN AACTOCTOC TATTITAGTA AACTOCTOCA CAAACTCCAG CTGGCCGTT TTTATGGCTT AAGGATATT AAGGATATT AAGGATATT TTTTGGCCTT TATTTCTTCT GGGAGGTAT AAGACTCCAG CTGGCCGTTA TGCTTTCTT ATTITGAGGAA CTCCATCTTC GGGAGGTA TAACACGCTA TAAAATTCC CTCATCTTC GCAACACTC TACACAGCTA TAAAATTCC TTTTGAAGAA ACACCTCCATT TTTGGTAATAA GCTCCATTCCT TTTTGAAGAA ACCTCCATTAT TTTGATAAAA GCTCCATTCAT TTTGTAATAA ACCTCCATTAT TTGGTAATAA GCTCCATTCAGCA ACTTCAGCA ACTTCAGCA ACTTCAGCA ACTTCAGCA ACTTCAGCA ACTTCAGCA ACTTCAGCA ACTTTGATTCATCAT ACCTCATCATA AAGCCAATT ATAGACATTC AGGTTATAAT ACTCATCACCA ACTTTTTTAAA GCTGCATTCATA AAGCCAATT ATAGACATTC AGGTTATAAT ACTCATCACCA ACTTTTTTAAA GCTGCATATTCATCACA ACTTTTTTAAA GCTGCATATTA AAGCCAATT ATAGACATTC AGGTTATAAT ACTTTATCACCA ACTTTTTTAAA GCTGCATATCA GCGGATATAAAAGCGCAT GAAGTGTATAAA GCTGCATATAA AAGCCAATTC AGGTTATAAT ACTTTATCACCA ACTTTTTTAAA GCGTGATTTCATCACA ACTTTTTAAA GCGAACCACT GAAGTGTATAAA GCGGCAAC GAAAGCCCAT GAAGTGCAT AAGACCAATC AGGTTATAAA GCTGAAATCA GCGGAACACAA GAAACCCAT GAAGTGTATTCAAAGCGAACAGAGGAA GAAACCCAT GAAGTGTATAAA AAGCCCAAT GAAGCCAAT GAAGCCCAT GAAGCCCAT GAAGTGCAAA GAACCCAAT GAAGCCCAT GAAGCCAT GAAGCCCAT GAAGCCAT GAAGCCCAT GAAGCCAT GAAGCCCAT GAAGCCCAT GAAGCCAT GAAGCCAT GAAGC			GTGCCAAGCT	TAAAAAaaac	manando	AMA A COMPCOM	CACCONCACC
ACTICATATA AGGATATTA CAGGAGATAT TATAGGCTAT TATAGGTATA AGGATATTA AGGATATTA CAGGAGATA TAGATATCATATC							
TTTATGGTTA AGGATATIC AGGAGGTA TTACCTATT CCATTATICGC ATATTCCTTCA ACAACTCCAG CTGGGCACA TGCTTCTTT ATTTAGCCC TTACTCTTC TGGGGGGTCA TCAACAGGTA TAAAATTCC TTTTGTAAGAA ACAACCTT TTTGGTAAA GCTCCCTTGC TAACAGGTGT TGTGAATACA ACAACCTT TTTGGTAAA AGTCCATCT CAAATACAGG CTAAATAATG ACCTGCATCA TCGCATCAGC ACATCACGC CTAAATAATG AATATCATCAT ACCTGCATCA TGGATAGAT AACTCACGC ACTTGGATTATTTTCTCT GGTCACTCAT TTGGTAAGA AGTCAACGC CTAAATAATG AATATCATCAT ACCTGCATTA TGGATAGAT AACTCACGC ACTTGGATTTTTCATCAT ACCTGCATTA TGGATAGAT AACTCACGC ACTTGGATTTTACACCA GCGGATATAAT TGGATAGAT AACTCACGC ACTTGTTGTATAAG GTTTTTTTT TAACACATTC AGTGTATAAT CCTTATCAAG GTTTTTTTTT TACACAATTC AGTGTATAAT CCTTATCAAG GTTTTTTTTT TACACAATTC AGTGTATAAT CCTTATCAAG GTTTTTTTT TACACAATTC AGTGTATAAT CCTTATCAAG GTTTTTTTTT TACACAATTC AGTGTATAAT CCTTATCAAG CTGGAATTCA GTTTTTTTTT TACACAATTC AGTGTATAAT CCTTATCAAG CTGGAATTCA GTTTTTTTTT TACACAATTC AGTGTATAAT CCTTATCAAG CTGGAATTCA GTTTTTTTTT TACACAATTC AGTGTATAAT CCTTATCAAG CTGGAATTCA GTTTTTTTTT TACACAATTC AGTGTATAAT CCTTATCAAG CTGGAATTCA GTTTTTTTTT TACACAATTC AGTGTATAAT CCTTATCAAG CTGGAATTCAAG GTTTTTTTTT TACACAATTC AGTGTATAAT CCTTATCAAG GTTTTTTTTT TACACAATTC AGTGTATAAT CCTTATCAAG GTTTTTTTTT TACACAATTCAACCC TCCTTGTTGC TCTTTTTAAG GTTGATTTAAGACATTC AGTGTATAAT ACTTAACACTCTC GTTTTTTTTTTTTTTTTTAAGACATTCAAGCCCTT GGGAATTCAAC GTTTTTTTTTTTTTTTTTTTTTTTTAAGACATTCAACCCTTTTTTAAGACATTCAACCCTTTTTTAAGACATTCAACCAATTCAACACCAATTCAACACCAATTCAACACCAATTCAACACAATTCAACACAATTCAACACAAAAAA							
ATTOCTICA ACACTCAG CTGGGCAGTA TGCTTTCTT ATCTTAGCAGA CTCATCTTTC TGGAGAGTCA TCAACAGCTA TAAAATTCC TTTGAACAGA CTCATCTTT TTTGGTAAAA GCTCCCTTG TAACAGGGT TGTGAATACA AACAACCTTT TTTGGTAAAA GCTCCCTTG TAACAGTGTG ATTTTCATCT GCTCCATCCA TCCAACTGC ACATACAGC CTAACATATATA ATTCATTA ACTCGCATTA TGGATAGAT ACTTCAGCA ACTTTTGAT TTTCATCTT TCTTGCTATA AGTTCCATAC TCCTTCTGC TCTTTTTAAG GTGGTTTTCA AAGCCAATCT ATAGACATTC AGTGTATAAT CCTTTTAAG CTGGATATCA gcgttacaag gggggggcac ttattacacaa agttttttat gttgagaata ttttttgat gggggggcac ttatttaga tctattttga tcgttcgctc aagAAGCGG CAGGGGTA GTTTTCTTT TCACCAGTMA GACGGGAAC AGAACGCCAT Gaggggctc ttatcatgca actcgtagga catgatgccgg cgcactgtgg gtgacacgggg ctgcacact accacqccg aacagggc cgcacact accacqccg accagcgc acagtccccg gccacgggg ctgcacact accacqccg acagtgccga catgtcccgg gcacacgggc ctgcacact accacqccg acagggccaca tagttccgg atctgcaca caggatgccg tttatcagg gatacatatg ctaactata ccgtaacac acgatacacg cttggggag cagaataga accgaacac gggtaacac tgcttccggt agcacactggc ctcagcatt tgagaagcac acggtacaca tgcttccggt accacatactg accacaacacg gggtaacac acgacacgc cctgccctga accacataac cggtaacac acaatacacg ccctgccgcacacaccacac	-						
TANCTICTIC TGGGGGTCA TCAACAGCTA TARAATTCCC TTTTGAAGAA CTCATCTTC CTTCTCCATC CAACCGCTT AGACAGGGT TGGAATACA AACAACCTTT TTTGTTAAAA GCTCCCTTGC TAACAATGT ATTTTTCTCT GCTCATCTTC TCCAACTGCA ACATCAACG CTAAATAATG AATATCAATTA ACCTGCTATA TTGGATAGATA ACCTGCATA CACTTGGAT TTTCATCTCT TCTTGCTATA AGTCCATAC TCCTTCTTGC TCTTTTTAAG GTAGTTTTTA TAGACATTC ATGACAATTC ACTCTCTTGC TCTTTTTAAG GTAGTTTTTA TAGACATTC ATGACAATTC ACTCTTATAAA CCTGCAATTCA GCGCATAC AGATCGCATA GGGGTATCA GGGGTCAA GAGGCCAA GAACGCCAT GGGAATTCA GGGGTCAACAGCAAT ATGACATTC ACTCTTTTTTTAAG GTAGTTTTTA TCACCTTC TCTTTTTTTAAG GTAGTTTTTA AAGCCAATTC ATGACAATTC ACTCTTATCAAG CGGAATCGCAT GGGAATTCA GGGGGTCAA GAGCGCAAC AGACGCCAAC AGACGCCAAC AGACGCCAAC AGACGCCAAC AGACGCCAAC GGGGGTGCGAACGCAAC AGACGCCAACCACACACA			TTTATGGTTA	AAGGATATTC	AAGGAAGTAT	TTAGCTATCT	CCATTATTGG
CTCATCTTTC CATCACTC CAAACCCGTT AAGACAGGGT TGTGAATACA AACAACCTTT TTGGTAAAA GCTCCTTGC TAAACAGGGT TGTGAATACA GCTCATCCC TCCAACTGCA ACATCAACGC CTAAATAATG AATATCATTA ACCTGCATTA TTGGATAGAT AACTTCAGCA ACCTTTGGAT TTTCATCCTC TCTTGCTATA AGTCCATAC TCCTCTTGG TGTTTTTAAG GTGGATTTTA AAGCCAATCT ATAGACATC ATGGATAAT CCTTATCAAG GTGGATTCA gcgttacaag tattacacaa agtttttta gttgagaata tttttttgat gggggccac ttattttga tcgttcgctc aaagAACCGG CGCCAGGGNT GTTTTTTTTT TCACCAGTNA ACTCAGTNA AGTCGAGAAC AGAACGCCAT Gaggggcctc attctctat ctgagtaca acgatcgca cagctgtcga gagcacccag gcacagggc ctgcaccat accacagccg aaacaagcgc ctgcaccat tatgttccgg atctgcaccat accacagccg aaacaagcgc ctgcaccat tatgttccgg atctgcaccat accacagccg aaacaagcgc ctgcaccat agtcaataaa cggtaaacc aggatagdt ctggctaccc tggtgaacac cctgcctga accgacaca acgaacagacc gcaaactagc cctagagcat tgagaagac acggtcacaca tggtcegg agacactagac gctaaacac agcaactagac dagaccagg cttcagagac cctgcctga accgagacc gggcaacaca agtccccag gcaaactagc cctgcacaca tatactac gcgtaaacac ggctaacaca tgcttcaggag agcatgaa accgagacc gggcaacaca agtccacac tcatgggg gcatgaga acctgagaca acggcacacac ttcatcaggag ggcatgag acctgaacac agcaactag ggcatgaacacac gcttaaacacacacacacacacacacacacacacacacac							
ACARCCTTT TTTGGTANAA GCTCCCTGC TAACATGGT ATTTTCCTCT GCTCATCC TCAACTGCA ACTCAACTGC ATAATAATA ATAATCATTA ACCTGCATTA TTGGATAGAT ACCTTCGTGC TCTTTGTAT TTTGATAGAT ACCTTCCTTCC TCTTTGTATA AAGCCAACTA ATAGCAATTC ATGACACTTC ATGATATAATA GTAACTTTTAA GAGCAATCA AAGCCAATCA ATGATAATAATA GTAACATTA AAGCCAACTA ATAGACACTTC ATGATATAATA CTTCTTTTTAA GTAGTATTAA GGGGATA AGGCAATCA GGGGATA AGAGCGAA AGGCAATA AGGCAATTA ATGACACTA AGGCAAA AGACGCAA AGGCACA AGGCACA AGGCACA AGGCACA AGGCACA AGGCACA AGGCACACACA							
GCTCATCC TCCAACTGCA ACATCAGC CTAAATATG AATATCATTA ACCTGCATTA TTGGATAGAT AACTTCAGCA ACCTTTGGAT TTTCATCTCT TCTTGCTATA AGTTCCATAC TCCTTCTTCT CTTTTTAAG GTAGTTTTTA AAGCCAACTT ATAGACATTC AGTTATAAT CCTTATCAGA gcgttacag tattacacaa agttttttat gtdgagaata ttttttttgat ggggcgcac ttatttttga tcgttcgcc aaagAACGG CGCCAGGGNT GTTTTTCTTT TCACCAGTMA GACGGCAAC AGAACGCCAT Gagggcctc attcttatt ctgagtaca acagtccgca cgcgtgtcgg gtagctcctt ccggtgggc ctgccaccat caccacgccg acaccatcactactgt attacagaag gcacacgccg ctgccaccat caccacgcgg ctgccaccat tatgttcgg atcgcacca cagtccccga gcaactagt ctacatactgt attacagaag gctaaccgt ttattacacgaa gcaacacgg ccaggacaca cctacactactg attacacgaag gcaacacgg tcagaacacag cctagacaca cctacactactg attacacgaag gcaacacgg gagaacacag cctagacaca accacagaga accagacaca accacagaga gcaacacgg cctagacaca cctacactacgt attacacta attacacgaa gcaacagac accagtacaca tatacacaaca gcctaacaca accacagaac accaacacacacacacacacac			CTCATCTTTC	CTTCTCCATC	CAAACCCGTT	AAGACAGGGT	TGTGAATACA
ACCTGCATTA TTGGATAGAT AACTTCAGCA ACCTTTTGAT TTTATACTCTC TCTTGTATA AAGCCAATC AGTTCCATAC TCCTTCTTGC TCTTTTTAAG GTAGATTTA AAGCCAATCT ATAGACATTC AGTTGATATAT CCTTATCAAG CTGGAATTCA gcgttacaag tattacacaca agttttttat gttgagaata ttttttga ggggcgcac tatttttga tggtagaata ttttttga gggggcgcac tatttttga tggtagaaca acagtccgca acagtgcccac accggtgggggggggg	<u> </u>		AACAACCTTT	TTTGGTAAAA	GCTCCCTTGC	TAACATGTGT	ATTTTTCTCT
TCTTGCTATA AGTTCCATAC TCCTTCTGC TCTTTTAAG GTAGTTTTTA AAGCCAATCT ATAGACATC AGTGTATAAT CCTTGATCAAG CTGGAATCCA gcgttacaag tattacacaa agtttttat gtgtagaata tttttttgat ggggcgcaca ttatttttga tcgttcgctc aaagAAGCGG CGCAGGGTA ATTTTCTTT TCACCAGTNA GACGGGCAAC AGAACGCAT Gageggcctc atttcttatt ctgagttaca acagtccgca ccgctgtccg gtagctcctt ttatcatgca actgtagga gcagtgcacaca cagtscccca gccacgggg ctgcaccat acagtscgcg cagcgccaca cagtcccccg ctcactctpt attacacgaag cctacactcgt tttatacagaag cagaataaat gatcatacg gcaaactggc ctcacactatg tgagaagcac acggtacacac ttttccggt agtcaataaa cagtaacac cctgccctga accgagaacc cctgccctga accgagaacc cctgccctga accgagaacc cctgcactta attacacta agcaataaac gcctaaaaa tactgttgta attacataag ggcatgatga acctgatgag ggcatgatga acctgatgag ggcatgatga acctgatgag ggcatgatga acctgatgag ggcatgatga acctgatgag ggcatgatga acctgatgag ggcatgatga acctgatgag ggcatgatga acctgatgag ggcatgtta aacaaaact ggcaacttgg tcgaaacacact gccatcatggt ggcacgttta aatcaaaact ggcaacacag cagcggcat atttaactt ggaaaaacata tctcaataa cgtaacacg tcgtggtata cacctggtg ggcaacacag cagcaccttg tcgaaacacac cagcaggac tcattgtgagaacac agcaccttg tcgcctacag ggcatgatga acctgaatcg ggcaaggtga acctgaacac ggcacgttta aacaaaact ggcaacttgg ggcaacacacac cctacacagac tcatcttgcg aaaacgggg gaaaactg ggcaacacacacacacacacacacacacacacacacac			GCTCCATCCC	TCCAACTGCA	ACATCAACGC	CTAAATAATG	AATATCATTA
AAGCCAATCT ATAGACATTC AGTGTATATAG CCTTATCAAG CTGGAATCG gcgttacaag tattacacaa agttttttat gttgagaata ttttttgat ggggcgccac tattttttgat tcgttcgctc aaagAAGCGG CGCCAGGNT CATTTCTTT TCACCASTNA CACGGCAAC AGAACGCCAT Gacggctct atttcttatt ctgattaca acagtccgca ccgctctgctg gtagctcctt ccggttgggc gggggcatga ctattcgtcg cgcacttatg actgctctt ccggttgggc gggggcatga ctatcgtcg cagcgccac acgctgtcccg gccacggggc ctgcaccat accacacgcg aaacaaggc cctgcaccat tatgttccgg atctgcatcg caggtgcgg cagcgccac acggtgccac cagaataat gacaatacg gctaaacgt ttttatcagg ctctggagg gcaaactggc ctcaggaatt tgaatacacg ggtaaacca tgttccggt gcaaactggc ctcaggaatc acgaacacg ggtgaaacac acggtcacac tcctcgct atatcactt accacacgcg acaacaaggcc cctcaccatctgt atatcactac accaacgcg acaacaaggcc acaacaagac acggacacaca cggtcacacat tgtggaaacac acggtcacac tgttccggt agcaactgga cccaggaacc acagaacacg ggtcaaacac tgttccggt accacgtta atacactaa accaacgagaat tgcttccggt tataatattt gccataaaa actacacacgc cgccctccac ccactcgcag gcaagaaagac acacaggca tactggtga acctgaatcg gaaaacacgg cagacacgg catcacagac ggcaagaacact tccaacagac ggcaagaacact ggcacacacag cacactttg gaaaaacacaca tccaagac gaaaaacacgg cacacactg cgcttacacacg cacactctgc gaaaacacgg cacacactg tggccttagaac acacagggaat ttccaacagac gagaaaacacacacacacacacacacacac	:		ACCTGCATTA	TTGGATAGAT	AACTTCAGCA	ACCTTTGGAT	TTTCATCCTC
gcgttacaag tattacacaa agttttttat gttgagaata ttttttttg ggggcgcac ttattttta tcgttcgctc aaagAAGCGC GCCCAGGNT GTTTTCTTT TCACCAGTNA atttcttatt ctgagttaca acagtccgca ccgctgtccg gtagctcctt ccggtgggcg cgggcatga ctatcgtcgc ccgcgtgtccg gtagctcctt ttatcatgca actcgtagga caggtcgcg acacagcgc cagcaccacat tatgttccgg atctgcaccat accacacgcg acacacagcg cctacactctg attacacaga gcgaacaggg ccgaataaat gatcatatacg ggaaactgg ctcaggaatc tcaattatta cctcacaggg gaagacctga gcaactagc ccaggaatc tcaattatta cctccacggg gaagacctga gcaactaga accggacaac acgacacagc acgacacagc acgacacagc cctgccctga accgacaac acgacacagc accacagcg accacagcg accacagcg accacaggg gaagacctga gcaactaaa ccggtaaacc acgacacaa acgacacagc tcatcacgct accacataact gccttaaaaa attatacact accacacagcg accacacacacacacacacacacacac			TCTTGCTATA	AGTTCCATAC	TCCTTCTTGC	TCTTTTTAAG	GTAGTTTTTA
gcgttacaag tattacacaa agttttttat gttgagaata ttttttttg ggggcgcac ttattttta tcgttcgctc aaagAAGCGC GCCCAGGNT GTTTTCTTT TCACCAGTNA atttcttatt ctgagttaca acagtccgca ccgctgtccg gtagctcctt ccggtgggcg cgggcatga ctatcgtcgc ccgcgtgtccg gtagctcctt ttatcatgca actcgtagga caggtcgcg acacagcgc cagcaccacat tatgttccgg atctgcaccat accacacgcg acacacagcg cctacactctg attacacaga gcgaacaggg ccgaataaat gatcatatacg ggaaactgg ctcaggaatc tcaattatta cctcacaggg gaagacctga gcaactagc ccaggaatc tcaattatta cctccacggg gaagacctga gcaactaga accggacaac acgacacagc acgacacagc acgacacagc cctgccctga accgacaac acgacacagc accacagcg accacagcg accacagcg accacaggg gaagacctga gcaactaaa ccggtaaacc acgacacaa acgacacagc tcatcacgct accacataact gccttaaaaa attatacact accacacagcg accacacacacacacacacacacacac			AAGCCAATCT	ATAGACATTC	AGTGTATAAT	CCTTATCAAG	CTGGAATTCa
GTTTTTCTTT TCACCAGTNA GACGGCAAC AGAACGCCAT Gaggegecte attetate totgagtaca acagtecque cogetitecy gtagctect teateateque coggetitecy acaccagt caggitectite teateateque coggetitecy acaccagt caggitectite teateateque coggetitect acactitity attaacqua accactaged aaccaacque citygaacac ctacatotity attaacquag coggetite teateateque coggetitectite agracaataa coggetite caggitectite teateateque cagataaca accactacaa coggetite des gagaacaca acggetite agracaataa coggetite agracaataa coggetite agracaataa coggetite attaacque gagaacque accactacate teatecque accactacate gectitaaaa attacque coggetite accactactite attaacatt attacactt attacactt attacagged accactaggite accactactite gaaaacata tectitaaaa attacque coggetite accactactite gaaaacact good aactagaac accactactite gaaaacactite gecatatatite gocactitie gocactitie gocactitie gaaaacacque cacactitie cacticacaga cactiticacacacque cacactitie cacticacaque accoctitie gaaaacacque cacactitie cacticacque accocque tacacque cacactitie cacticacque accocque tacacque cacactitie gaaaacacque cacactitie cacticacque acciticacacacque cacacticie daacticacacacque acacacacacacacacacacacac			gcgttacaag	tattacacaa	agtttttat	gttgagaata	tttttttgat
SEQ ID: 19 PACKO- Alt84TGA SEQ ID: 19 PACK			ggggcgccac	ttatttttqa	tcgttcgctc	aaagAAGCGG	CGCCAGGGNT
attictiati cigagitaca acagicoga cogeticog glagototti coggigogogogogogogogogogogogogogogogogog	1		Ch.h.h.h.h.C.h.h.h.	TCACCAGTNA	GACGGGCAAC	AGAACGCCAT	Gageggeete
ccggtgggcg cggggcatga ctategtege cgcaettatg actgtettet ttateatga actcgtagga caggtgecgg caggeceaa cagtececeg gcaegggce tggcaecat accaegggg cagggtggg cagggtggg ctggcaecat tatgtteegg atetgeateg caggatgetg ctggetaece tgtggaagg cagataaat gateatateg beaattata cetecaeggg gagagcetgg agtaaacegg ctaaggatg cteteggagg gcaaactgg ctaggataac agcaataga atgtaatateg cetgecetga accgaegaeggggaatt tgetteeggagg gagagcetga agcaatagae atataeggegt accaataaat gcettaaaaa aattaeggee cgceetgeea accaataact geettaaaaa aattaeggee cgceetgeea accaataact geettaaaaa aattaeggee cgceetgeea tataatattt geeatggg aaaacggggggaatt tgettetgaa tteatgggg tataatattt geeatggtg accaggggat cagaagagg cgaaggggggaaaacaggggggaaaacaggggggaaacagggggaaacagggggg			atttcttatt	ctgagttaca	acagtccgca	ccactatcca	gtageteett
ttatcatgca actogtagga caggtgcogg cagcgccaa cagtccccgg gcacagggc ctgcaccat acccacgcg aaacaaggc ctgcaccat tatyttcogg atotgcacat cagatyctg ctggctaccc ttgtgacacat cagatytcogg atotgcacat tatytcogg atotgcacat ttgagaaggc cttggaggg cagaataaat gatcatatag tcaattatta cetccacggg gaagactga agcaataaa ccggtaaacc aggtcaaac tgcttcoggt agtcaataaa ccggtaaacc aggtcaaat tcatccgct attatacact tatatacact gccttaaaaa aattacgcc accaggat ttcatcacgat accaataact gccttaaaaa aattacgcc acaggaggac accaggc ggcatatatt ggcaatatatt ggcacactgg ggcatgaga actgagaag catcacagga ggcatgata acctgatgga accaggggat accagacacgg ggcatgata accagagac ggcatgata acctgatga ggcacactgg ggcatgata acctgatgg aaaacggggg cagaagaatt ggcacattt ggcacgttta aatcaaaact ggtgaaact accaggagt ggcacatttg ggcacgttta attcatatat ggcacgttaaaaa acctttagg gaaataggc aggtttaca cgtaacacg cacatcttgc gaaatatatgt gtagaaact accgggat tcgtaacacg cacatcttgc gaaatatatgt gtagaaact accgggat tcattgcat acgggatat acacggggat acacgggga cagagaatg tcattgcat acggagata cacatctcca acggagat tcattgcat acggagata cacataccag cgtttaatt ttcttaagg gaaataggc aggtttaa cacgggttaa caagggtgaa cactaccag ttcattgcat acggagataca acggtcaga cggtagaaacggagaaaaa atacgccgg taggaaactg tcattataa accttaaaaa ttcaccaga cggtcagat ttcattgcat acggagatatt ttcattacaa acctatccaa tatcaccagc cggaaatcg tcacaggatat ttcattacaaaa ttcaccaga cgatgaaaacg gtttcagtt gctaatgaa cacggtgaa cactaccaga ttattacacag ccggaaatggaa acctaacaga cggtctggt ataggaaact tatcaccagc tcacaggat tttttcacgat gcaattgga ataaaacggggaaacctgaaaccag gattaatta ttacacaag ggaacctaca acggaccaaggccaaggcctgaaaccaaggattattattacaaaa acggacacaaggattataa tttcacaaaaa atacgcccgg taggatact ttttcacgaa acggacacaaggct caccaaaaaa tacgcccgg taggatact ttttcacgaa acggacacaaggccaaaggctcaaaaaaaaaa			ccaatagaca	caaaacataa	ctatcatcac	cacacttata	actgtcttct
gccacgggc ctgcaccat accacgccg aaacaagcgc cctgcaccat tatgttcegg atctgcatcg caggatgctg ctgggacacc ctacatctgt attaacgaag cgctaaccgt ttttatcagg cagaatagac gcaaatagac attaacacgag gcaaacatggc accaggaatgg gcaaacatggc accaggaatgg gcaaacatggc accaggaatgga accaggac cctgcactga accagacac accagacac accadacacac accadacacac accadacacac accadacacac accadacacac accadacacac accadacacac accadacacac accadacacac accadacacacac			ttatcatcca	actcotagga	caggtgccgg	cagcgcccaa	cagteceeeg
tatgttccgg atctgcatcg caggatgctg ctggctaccc tgtggaacac ctacactctgt attaacgaag cgctaaccgt ttttatcagg cagaataaat gatcatatcg tcaatatta ccctcacagg gaagacctga gcaaactggc ctcaagcatt tgagaagaca acggtcacac agcaataaac cagcaataaac cagcaataaac cagcaataaac cactgccctga accaataacc ggctaaacac attatacggct attaacgac tcatcgctt attatacactt attacaggcg accaaggc ggctgaatt tgctttcgaa tttctgccat tactgttgta attaatactt gccattgtga accaagacggcat cagcaccttg gccatgata acctgatgat gccatgggg cacagggggaaacaca aaatagggcc ctcatcagag gaaaaacata ttctcaataa acctttaggg gaaagaacgt tgcttcagg tataatattt gccacttgg gaaaacggggg cgaagaagtt gccatattg gccactgtta accaaaact ggtgaaactc accaagggat tggctgagac gaaaaacaca cacaattgg gaaaacacg gaaaaacaca ttcatgccag cacatcttagg gaaataaggcc aggtttcac gaaaacac cacaacttgg gaaatatatgt gtagaaaccg cacaacttg gaaaaacac ggtgaaaacc gaaacacgg cacaacttg caggagaa acggaaacacg cacaacttg gaaaaacacg gtttcaataa acctttagg gaaaacacg cacaacttg gaaaaaccg gattcaatggaa acggaaacacg cacaacttacca tacacaagga tcacaggaaacggaacacggaaacacgaacacgaacacgaacacgaacacgaacacgaacacgaacacgaacacgaacacgaacacgaacacgaacacgaacacgaacacgaacacgaacacgaacacacacacacacacacacacacacacacacacacaca			accacacaca	ctcccaccat	accracacca	aaacaagcgc	cctgcaccat
ctacatctgt attaacgaag cgctaaccgt tttttacagg ctctggagg cagaataaat gatcatatcg tcagaattaata gatcatatcg gcaaactgge ctcaggagt tegaagacca acggtcaaca tgcttcoggt agtcaataaa ccggtaaacc agcaatagac acggtcaaca tgcttcoggt agtcaataaa ccggtaaacc agcaatagac attaacgcc ctgccctga accagacgac gggtgaatt tcatccgctt attatcactt attcaggcgt accagaggg gtttaagggc accagagac tacttgtgta attaatattt gcccattggtga aaattcaggcg cacatgggg ggaagatgatg acctgaateg cacagacactgg ctacacagac ggcatgatga acctgaateg cacaggaat cagaacacttg tcgcttgcg attaatattt gcccattggtga aaaacggggg cagaagaagtt gtcatattt ggcaatggta accagaacac ggtagaacact ggaaaaacata ttctcaataa accctttagg gaaaaacata tcgtgggaa cacatctgg gaaaaacata cgaaaacaca acccttagg gaaaaacata tcgtgggaa cacatctgg gaaaaacata cgaaaacaca acccttagag gaaaaacata tcgtgggaa cacacatcg gaaaaacac ggaaaaacac ggaaaaacggggaa cacacatggaac cacatctgc gaaatatgg tcgtggaaacc cgagggtgaa cacagggggaa aagaatgga ggaaaacag ggaaaacag ggaaaacag ggaaaacag ggaaaacag ggaaaacag ggaaaacag cacacttg cgaaaacag cacatctcca acggggtaa aacggtgaa cacaggggga aagaatgga ataaaggccg tcactggaa tacacagac tcaccggctt ttcattacaca ggaaaacac ggaaaacag gaaaacag ggaaaacag cacacagc tcaccgtctt tcattacaca acgggttaa tacacacag tcaccgtctt tcattacaca acgggttat ttcttacaga gccattggga ataaacacg ggaaaacacg cagacacagg cacacaggc tcaccgtctt tcattacaca acggggtaa accgggaaacacag gaaaacaggc accagggaat tacacacaga cacacatcacaaaaca ggacacagg cacacaggc tcaccggaatacaacaggaaacacaggaaacaggcacaaggcacacaggaacacaggaacacaggaacacaggaacacaggaacacaggaacacaggaacacaggaacacaggaacacaggaacacaggaacacaggaacacaggaacacaggaacacaggaacacaggaacacaggaacacaggaacacaggaacacaaaacacacacacacacacacacacacacacacac			tatattaga	atetecatee	carratecte	ctggctaccc	tatagaacac
cagaataaat gatcatatcg tcaattatta cotocacgg gagagcotga gaaactggc agtcaataaa cocgataaacc agcaatagac ataaacggct atttaacgac cctgccctga accgacgacc aggatcaatta ttottcagat tcatcagtta atcattaag cattatgcgc acatcagag ggcatgatta atcattaggcg tactatatatt gccatcagtg attcattaag cattatgcgc cgccctgca ctcatcagag ggcatgatga acctgatgg attcattaag cattatgcgc cgccctgca ctcatcagag ggcatgatga acctgatgg acctgatggg acatgatgg acctgatggg attaatattt gccatggtg aaaacggggg catcacagac ggcatgatga acctgatgg aaacggggggggaaacac cagcaccttg tgccttggg gaaaacata ttcaaaaac ggtgaaactc caccaggggat tgccatattg gaaaacata ttcaaaaac ggtgaaacca caccaggggat tgccatattg gaaaacata tcatcaaaa accctttagg gaaataaggc aggtttcac cagtggatt cactcagag ggatatatgt gtcaatggaa cactatggaa aacgggtgaa cactaccaa ggtgaaactg caggaaactg tcattggaa acatcaggggga aaggattga cactatggaa cactatggaa tactacaaact tcattgcca acggcttat ttcttacgat gaaaactt tcattgcaa cggttgatt ttctttacgat cagtgattt ttcttcaat taggtcaat tggcgatata tggcgaatta tcagtgattt ttctccatt taggtcact taggtactt ttcttacat ggaaactt gggaacctct aacgcacgggaacccag gattattat ttctggaagggaacccag ggattatat ttctggaagggaacccag ggatatatca acgggcttca tggaacccag ggatatatcaggaagggaacccag ggaaaccag ggatatatca ttcagggggaacccag ggaaaccag ggatatatcaggaaggaccaggaaccag ggaaaccag ggatatatcaggaaggaccaggaaccaggaaccag ggatatatcaggaaggaccaggaaccag ggatatatcaggaaggaccaggaaccag ggatatatcaggaaggaccaggaaccag ggatatatcaggaaggaccaggaaccag ggatatatcaggaaggaccaggaaccag ggatatatcaggaaggaccaggaaccag ggatatattatagggaaccaggaaccag ggatatattatagggaaccaggaaccag ggatatattatagggaaccaggaaccag ggatatattatagggaaccaggaacacaggaaccaggaacacacagacaggaacacacagacacaggaacacacagacagacacacagacagacacacagacacacacagac	ļ		catgetetet	atteaccasc	cactaaccat	ttttatcacc	ctctgggaga
geaactgge etcaggeatt tgagaagca aeggteacac tgetteeggt agteataaac ecggtaaacc ageaatagac tetettegaa tttateageat teatecgett attateacett attateacett teateggeggge accaataact geettaaaaa atteatgee eaceagggat tattaaggge accaagac ggeatgata atteattaag eactgaagge ggeaagacgggggatgatatteataatatt geeaagaggggaactgggaactgggaaaacaaggggggaaaacaagggggggg			clacatetyt	actaacyaay	tgaattatta	cctccaccagg	gagaggggg
agteaataaa ceggtaaace ageaatagae ataageget atttaacgae cetgeectga acegacgaee gggtegaatt tgettegaa tttetgeeat teatecgett attateactt atteageget ageaceageg gtttaaggge acegacatagae teategttga aceaataaaet geettaaaaa aattaegeee cegeectgeea etacagagagaege ggeatgatga acetgaateg caagegggag caacacettg tegeettgeg aateaaaet ggeeatggg aaaaeggggg gaagaagagt ggeatgaga eegaaaaee ggaaaaaeet ggeaaaaee ggaaaaaeet ggeaaaaee ggaaaaaeet ggaaaaeee gaaaaaeeg eacettage gaaatatagt ggaaaaeeg eacetgggaateeggaaeaeaegggaaeaeaeggagga			Cagaataaat	gattatatty	tanganggag	accetacggg	tacttccaat
cctgccctga accgacgacc gggtcgaatt tgctttcgaa tttctgccat tcatccgctt attatcactt attcaggcgt agcaccaggc accadaact gcttaaaaaa aatacgccc cgccctgcca tactgttgta attcattaag cattctgcg acatggaagc ctcatcgag ggaatgatga acctgaatcg ccagcggcat cagcaccttg tataatattt gcccatggtg aaaacggggg cgaagaagtt ggccacgtta aatcaaaact ggtgaaacct cagcaggtat tggctgagac gaaaaacata ttctcaaaaa ccctttagg gaaataaggcc cggaaaacag tcgtgggatt cagtgggat cagtacaggc cacatcttgc gaatatatgt gtagaaact ccggaaatcg tcgtggtatt cactccagag cgatgaaaac gtttcagtt tcattgccat acg gaaataaggcc gagaatatgg tcgtggtat cactccagag cactatccca tatcaccag tccatggaa aacggtgtaa cactatccca tatcaccag tccatggaa tcattgccat acg gataaaaactt tccattgccat acg SEQ ID: 19 packo- A184TGA packo- A184TGA gaactccgga tgagcattca tcaggcgggc aagaatgtga ggcgtaata tccagctgaa cggtctggtt ataggtacat tgagcaactg ggcgtaata cagtgtattt ttctccaat ttagctcct tagctcctga aactgaaatgc ctcaaaaagt tctttacgat gccattggga tatatacacg gtggtaatc cagtgatttt ttctccatt ttagcttct tagctcctga aactcaaaag tggacctct ttagctcct tagctcctga aagttggcc tgagcactat attacacag ggaacctct aggtgccat ttagcacat ttttcgcaa ggagcaccag ggttattt ttctgcgaag tgagcaccag ggttattt ttctgcgaag tgagcaccag ggttattt ttctcgaag tgagcaccag ggttattt ttctgcgaag tgagcaccag ggttattt ttctgcgaag tgagcaccag ggttattt ttctgcgaag tgagcaccag ggttattt ttctgcgaag tgagcaccag ggttattt ttctgcaaa aagttggcc tgagcaacta ctgatttagt gtagaacgg caaaagtgcg tcggagaaccacta ctgatttagt gtataatgg gtttttgag ggttactcg tcacaggat tcatcccgtg tccaccatgg ggcttccgtt gggtaaccac gggcacaca gggtggtggt gggtaacgg ggctacacag ggctacacag ggctacactag gggttacact ctccatgtg aagtggttgcacaggaaaaa aaggctgcac aaaggggttg aggaaatag cgttcactggaa aaggctgcac aaaggaggtg cagtgaacactg gggaacaccag gacgaacatg ttcattggg aataacac gggggtggt aggaaacacag gacgaaaaaggccg caggaaaaacacag aaggaacacag gacgaacacag gacgaacacag aggtgatacacaggacacagggacacacaggacacacag gacgaacacag aaggtgacacacag aaggtgacacacag aaggtgacacacag aaggtgacacacag aaggtgacacacag aaggtgacacacag aaggtgacacacag aaggtgacacacag aaggtgacacacag aaggtgacacacacag aaggtgacacacacag			geaaactgge	cccaggeact	cgagaagcac	ataaccact	atttaaccac
tcatccgctt attatcactt attcaggcgt agcaccaggc gtttaagggc accataact gccttaaaaa aattacgccc cgcctgcca tactgttgta attcattaag cattctgccg acatggaagc ggaatgatga acctgaatcg ccagcggcat cagcaccttg tcgccttgcg tataatattt gcccatggtg aaaacggggg cgaagaagt ggcacgttta aatcaaaact ggtgaaactc accagggat ggcacggtta aaccaaact ggtgaaactc gaaaaacata ttctcaataa acctttagg gaaataggcc cggaaaccg cacatcttgc gaatatatgt gtagaaactg tcgtggtatt cactccagag ggtgaaacac gtttcagttt cactgggtat cactccagag ggtgaaaca gtttcagttt tcattgccat acg ggatgaaaca gtttcagttt tcattgccat acg ggatgaaaca gtttcagttt tcattgccat acg ggatgaaaca ttcttaaaaa acctttacca tatcaccagc ccggaaatcg gccattggaa acagggtgaa cactatccca tatcaccagc ccggaaatcg gccattggaa acagggtgaa cactatccca tatcaccagc tcaccgtctt tcattgccat acg ggatgaacac ggtttaatt ttctttaacgg tctttaaaaaa gggccgaatta tccagctgaa cggttattt ttcttacgg tcttaaaaaa gggccgaatta tcaacaaaa tacgccgg tagtgatcat taggcaactg acggaaattg ggaacctct acgggaactct taggtactt tattcattat ggtgaaagtt ggaacctct acgggacctca agggctccc ggatcaaca gggaccacaa gggaccacaa agggcttccc ggatcaaca gggacaccag gatttattt ttctcgcaaa tgggctacaactg tggtactata ttattcggcg caaagtggcc caacgtctca tttcgccaa aagttggccc tagacctta ctgattagt gtatgatggt gtttttatt ttctgcgaag tggtatatta ttctcggcg caaagtggcc caacgtctca ttattcattat ggtgcaactg ggaccacctg gattatatt ttctcgcgaag tggtatatct ttctcgcaactg ggatcactgt caacgtctca ttattcattat ggtgcaactg ggatcaccactg ggatcacactg ggatcaccactg ggatcacactg ggatcacactg ggatcacacacacacacacacacacacacacacacacaca			agtcaataaa	ccggtaaacc	agcaacagac	tactttcaaa	tttctcccat
accaataact tactyttyta atteattaag cattetgeeg acatygaage cateacagae ggeatgatga acctgaateg ceageggeat cageacetty tegeettgeeg tataatattt geceatgytg aaaaeggggg cgaagaagtt gteetattty gecaatgtg aateaaaact ggtgaaacte accagggat tggetgagae gaaaaacata tetetaataa accetttagg gaaataaggee ceggaategg teggtgatt aaeggtgtaa cacteagag gaaataatgt gtagaaacte ceggaategg teggtgata cacteagag cgatgaaaac gttteagttt teattgeeat acggtgtaa cactatecea tateaceage teaecggtta aaeggtgtaa cactatecea tateaceage teaecggttt teattgeeat acggtetgat teettaaga ggeetaata teaecage teaecggttt teettaaaaa ggeetggt ataaaactt teettaagg teettaatat teettaaga ggeetaata teeteaga cagtattat teettaaga ggeetaata teaecage teettaaga ggeetaata teaecage teettaaga ggeetaata teaecage teettaagaa ggeetaata teaecage teettaaaaa ggeetagat teettaaga ggeetaata teaecagaa teettaaga ggeetaata acggegtate teettaagaa ggeetaata teaecagaa teettaaga ggeataate teaecagaa actgaatte teettaaga ggeatetgg teettaaaaa ggeetaata acteaaaaa atacgeegg tagtgatet tatetaaacg gtggtatate cagtgattt teettacgat tagtgatett tatetatat ggtgaaagtt ggaacetett aggeeteegg tagtgatet atteteatat ggtgaaagte tagtateteegg teaecaggaat tateteage gggtaacea ggattatta teetgegaag tegeaacea gattattat teetgegaag tegeaacea ggeetaatag geetteegtt geetaacaa ggacacaa gattatta teetgegaag tegeaacea geetteeggt tegeaacea ggeetaacea gggtatee tageacea gggtatee tageacea gggtatee gggtaacea aaggeetee teaeaggat teaeacagga geetaacea ggeetaacea aaggeetee teaeagaagaa aaggeetee caggagagaa aaggeetee teaeagaacea caggagaacea cateaceaca ggeetaacea ggeetaacea ggeetaacea ggeetaacea ggeetaacea ggeetaacea ggeetaacea aaggagtgaa caggagaaaaa aaggeetee teaeagagagte cagagaaatag cgteagaac		'	cetgeeetga	accgacgacc	gggtcgaact	agaagagag	atttaaaaa
tactgttgta attcattaag cattctgccg acatggaagc catcacagac ggcatgatga acctgaatcg cagcggcat cagcaccttg tcgccttgcg tataatattt gcccatggtg aaaacggggc cagaagagtt tggctagagc ggaaaacata ttctcaataa accctttagg gaaatatgg tggctagagc cgtacaccgc cacatcttgc gaatatatgt gtagaaactg cggtagaaccg cggaaaacgg tggctagagc aggatttcaccgtacacgc cacatcttgc gaatatatgt gtagaaactg gtccatggaa cactggggtat cactccagag gattatacg gtagaaaccg ccggaaatcg tcgtggtat cactccagag cactatccca tatcaccagc tcactggaa taaaagggcg gttcatggaa cactatccca tatcaccagc tcactggaa tcactggaa taaaaggccg ggtagaaacct tcacaggtgaa cactatcca tatcaccagc tcactggaa tcacaggtgaa cactatcca tatcaccagc tcactggaa tcacaggtgaa cactatcca tatcaccagc tcacaggtat tcattacaga tcacagaactg ggccgtaata taaggtacat tgagcaactg ggcgtaata taaggtagaactg cagtggattt ttcttacgat taaggtacat taggcaactg ggggaactct taggtgattt ttcattat ggggaactct taggtgatct taggtactt tattcattat ggggaacttca aagggcttccc ggtatcaaca gggacaccag gatttatta ttctgggaagtt tgatcacaa agggcttccc ggtatcaaca gggacaccag gatttatta ttctgggaagtt tgaccactta ctgatttagt gtatgatggt gtttttgagg tggtagacgggtagct gggtaacggc gggtaacgc ccggaacaca ggcgtaggggtggt caagggaaaa aaggctgcac ccggacatca gcgctagcgg gggtatact cattattccg tcacaggta tatacacag ggttacacc gggtacacca gggtaacgc gggtaacgc caaagggtgt cagtgaactg tcactctgt tcacaggata tcacaggat tatacacag ggttatact tcacaggta ttatcactga gggtaacgc gggtaacgc gggtaacgc caaaggtgcg caggagaaa aaggctgcac ccggacaca gcgctagcgg agtgatacc caggagaaaa aaggctgcac cgggacacca ggcgtaacgg ggttacacagga tatattccg tcactgcg gagaaatat ggtacacagga tatattccg tcactgtg ggcgaaaacg cggtagacgc cagagaatat gggctacacaggat tatattcacaga ggctacacaggat tatattcacaga acgggtagacacag ggctacacag ggcgaaacacag ggctacacag ggcgatacacag ggctacacag ggcgaaacacag ggcgaaacacag ggcgatacacag ggcgatacacag ggctacacag ggcgatacacag acggatacacag ggctacacag ggctacacag ggctacacag ggctacacag ggctacacag ggctacacag ggctacacag ggctacacag ggctacacag acggatacacag acggctacacag ggctacacag acggatacacag ggctacacag ggctacacag ggctacacag ggctacacag ggctacacag acggatacacag acggctacacag ggctacaca			tcatccgctt	attateactt	acteaggege	ageaceagge	ctcatcacac
ggcatgatga acctgaatcg ccagcggcat cagcaccttg tegecttgeg tataatattt gccacgttta aatcaaaact ggtgaaactc accagggat tggctgaac agaaaacata ttctcaataa accctttagg gaaataggc cggaaaactg ccgtaacacgc cacatcttge gaatatatgt gtagaaactg ccggaaatcg tcgtggtat cactcagag cgatgaaaac gtttcagt tcattgccat tcattgccat acg gaatatatgt gtagaaactg ccggaaatcg ccggaaatcg cacatctcca acg gatgaaaac gtttcagttt aacgggggaa cacatatccca tatcaccag tcattggaa tcattgccat tcattacaca ggccgtaata accgacgtgatttt tcattacgat gccattggga tatatcaaca gtggtaatatc taggtacctt tagctccta acgtccaa acggctcca ttttcattat ggtgaaagtt aggcacctgat tcattccat ttatcattat ggtgaaagtt aggcacctgat tcattcattat gggaacctct agggacccacag gattattta ttctcatat tggtaacctg tcacaggtat ttattcggcg caaagtgcgt tggcaacgg ggctaacac gggataccacag gattattta ttctgcgaag tggctaccaggaagt tcattagggatgt tcattaggg caggaaatag cggggtggt tggctacatat gggtaacacca gggctaacaca agggttgt tcattaggg caggaaatag tcattgggatgt tcattggga aagggatga agggtaacctg agggtaacccag ggctaacaa aggggtggt caacgggagaa acccattgaggaacacacacacacacacacacacacacac]		accaataact	gccttaaaaa	aattacgccc	egegetgeta	citaticgcag
tataatattt gcccatggtg aaaacggggg cgaagaagtt gccatattg gccacgttta aatcaaaact ggtgaaactc acccagggat tggctgagac aggaaaacacacacacacacacacacacacacac			tactgttgta	attcattaag	cattetgeeg	acacggaage	tagaattaga
SEQ ID: 19 pACKO- A184TGA gaactccgga tcttacgaaactt tctcaataa accctttacgaaatggaacttcaaaaactt tccagaaactt tccagaagaacttaaaactt tccaaaaactt tccagaagaactaaacggtgaaacgaaac			ggcatgatga	acctgaatcg	ccagcggcat	cagcaccity	ctgccttgcg
gaaaaacata ttctcaataa accetttagg gaaataggee aggttteac egtaacaege tegtggtatt cactecagag eactatetee gaatatatgt gtagaaactg eegtaatgaa aacggtgtaa eacggetgaa eacggetgaa eacggetgaa eacggetgaa eacggetgaa eacggetgaa eacggetgaa eacggetgaa eacggetgaa eacggetgaa eacggetgaa eacggetgaa eacggetgaa eacggetgaa eacggetgaa eacggetgaa eacggetgaa eacggetgaa eacggetgaata eacggeteea eacggeteea eacggeteea eacggeteea eacggeteea eacggeteea eacggeteea eacggetgaata eacggagaataa eacggetgaata eacggetgaata eacggagaataa eacggetgaata eacggagaataa eacggagaataa eacggetgaataa eacggagaataa eacggagaataa eacggagaataa eacggagaataa eacggagaataa eacggagaataa eacacggaataataa eacgceggaataa eacgaataa eacggagaataa eacacggaataataa eacggagaataa eacacggaataataa eacgceggaataaa eacacggaataaa eacacggaataaa eacacggaataaaaaaaa eacacggaataaaaaaaaaaaaaaaaaaaaaaaaaaaaaa	1		tataatattt	gcccatggtg	aaaacggggg	cgaagaagtt	gtccatatty
Cgtaacacgc cacatettge gaatatatgt gtagaaactg ceggaaateg tegtggtatt cactecagag egatgaaac gttteagttt teattgecat acagggtgaa cactatecca tateaceage teacegtett teattgecat acg SEQ ID: 19 pACKO- A184TGA gaacteega tgagcattea teaggegge aagaatgtga ataaaggeeg gataaaactt gtgettattt teetttaegg tetttaaaaa ggeegtaata teeageegga eggettaett teeteegat teaggeaateg eeggeaateg eeggeegtaata teageegga teagtgatett eeggeaateg eeggeegtaata eacteaaaaa ataegeegg taggateet tageteetga aaatetegat ggaacetett eeggeegat eaaeggeetee eagggetteee eaggeettee eagggetteee ggaaceeg eaaeggeetee eeggaaagte eeggeegat eeggaagteeggat eeggaagteeggat eeggaagteeggat			gccacgttta	aatcaaaact	ggtgaaactc	acccagggat	tggetgagae
SEQ ID: 19 pACKO- A184TGA gaactccgga tgagcattca tcaggcgggc aagaatgtga ataaaggccg gataaaactt tctttacgat tcaggtgtt tccaggatgatatatcaggtgtt ttcttacagt tccaggatgatatatcaggtgtaactcaggtgtaactggt tctttaaaaa ggccgtaatatcaggtgtt tttttacgat tcaggtggattt tttttacgat taaggtacat tgagcaactg gtggtatatc cagtgatttt tttttccatt ttagcttct tagctcctgat aactccgat aactcaaaaa atacgccgg tagtgatctt atttcattat ggtgaaagtt ggaacctct aggacctcta ttagctccta ttttcgcaa aagttggccc agggcttccc tagtatatatattcggg ttatattata			gaaaaacata	ttctcaataa	accetttagg	gaaataggcc	aggttttcac
SEQ ID: 19 pACKO- A184TGA gaactccgga tgagcattca tcatgggggc aagaatgtga ataaaggccg gataaaactt gtgettattt teettacgg tcatatacaacaggggggggggggggggggggg	1		cgtaacacgc	cacatcttgc	gaatatatgt	gtagaaactg	ccggaaatcg
SEQ ID: 19 pACKO- A184TGA gaactccgga tgagcattca tcaggcgggc aagaatgtga ataaaggccg gataaaactt gtgcttattt ttctttacgg tcttaaaaa ggccgtaata tccagctgaa cggtctggtt ataggtacat tgagcaactg actgaaatgc ctcaaaatgt tctttacgat gccattggga tatatcaacg gtggtatatc cagtgatttt tttctccatt ttagcttcct tagctcctga aaatctcgat gaacctctt acgtgccggt tagtgatctt atttcattat ggtgaaagtt ggaacctctt acgtgccgat caacgtctca ttttcgccaa aagttggccc agggcttccc ggtatcaaca gggacaccag gatttatta ttctgcgaag tgatcttccg tcacaggtat ttattcggcg caaagtgcgt cgggtgatgc tgccaactta ctgatttagt gtatgatggt gcttctgttt ctatcagctg tccctcctgt tcagctactg acggggtggt gcgtaacggc aaaagcaccg ccggacatca gcgctagcgg agtgtatact ggcttactat gttggcactg atgaggtgt cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggtgcgtca gcagaatatg tgatacagga tatattccgc ttcctcgctc actgactcg tagccctggt cggttcgactg			tcgtggtatt	cactccagag	cgatgaaaac	gtttcagttt	gctcatggaa
SEQ ID: 19 pACKO- A184TGA gaactccgga tgagcattca tcaggcggc aagaatgtga ataaaggccg gataaaactt gtgcttattt ttctttacgg tctttaaaaa ggccgtaata tccagctgaa cggtctggtt ataggtacat tgagcaactg actgaaatgc ctcaaaatgt tctttacgat gccattggga tatatcaacg gtggtatatc cagtgatttt tttctccatt ttagctcct tagctcctga aactcaaaaa atacgccgg tagtgatctt atttcattat ggtgaaagtt ggaacctctt acgtgccgat caacgtctca ttttcgcaa aagttggccc agggcttccc ggtatcaaca gggacaccag gatttatta ttctgcgaag tgatcttccg tcacaggtat ttattcggcg caaagtgcgt cgggtgatgc tgccaactta ctgatttagt gtatgatggt gtttttgagg tgctccagtg gcttctgttt ctatcagctg tccctcctgt tcagctactg acggggtggt gcgtaacggc aaaagcaccg ccggacatca gcgctaactg agtgtatact ggctactat gttggcact atgagggtgt cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggtgcgtca gcagaatatg tgatacagga tatattccgc ttcctcgctc actgactcgc tacgctcggt cgttcgactg			aacggtgtaa	caagggtgaa	cactatccca	tatcaccagc	tcaccgtctt
A184TGA gataaaactt gtgcttattt ttctttacgg tctttaaaaa ggccgtaata tccagctgaa cggtctggtt ataggtacat tgagcaactg actgaaatgc ctcaaaatgt tctttacgat gccattggga tatatcaacg gtggtatatc cagtgattt tttctccatt ttagcttcct tagctcctga aaatctcgat aactcaaaaa atacgcccgg tagtgatctt atttcattat ggtgaaagtt ggaacctctt acgtgccgat caacgtctca ttttcgccaa aagttggccc agggcttccc ggtatcaaca gggacaccag gatttattta ttctgcgaag tgatcttccg tcacaggtat ttattcggcg caaagtgcgt cgggtgatgc tgccaactta ctgatttagt gtatgatggt gtttttgagg tgctccagtg gcttctgtt ctatcagctg tccctcctgt tcagctactg acggggtggt gcgtaacggc aaaagcaccg ccggacatca gcgctagcgg agtgtatact ggcttactat gttggcactg atgagggtgt cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggtgcgtca gcagaatatg tgatacagga tatattccgc ttcctcgctc actgactcg tacgctcggt cgttcgactg			tcattgccat	acg			
A184TGA gataaaactt gtgcttattt ttctttacgg tctttaaaaa ggccgtaata tccagctgaa cggtctggtt ataggtacat tgagcaactg actgaaatgc ctcaaaatgt tctttacgat gccattggga tatatcaacg gtggtatatc cagtgattt tttctccatt ttagcttcct tagctcctga aaatctcgat aactcaaaaa atacgcccgg tagtgatctt atttcattat ggtgaaagtt ggaacctctt acgtgccgat caacgtctca ttttcgccaa aagttggccc agggcttccc ggtatcaaca gggacaccag gatttattta ttctgcgaag tgatcttccg tcacaggtat ttattcggcg caaagtgcgt cgggtgatgc tgccaactta ctgatttagt gtatgatggt gtttttgagg tgctccagtg gcttctgtt ctatcagctg tccctcctgt tcagctactg acggggtggt gcgtaacggc aaaagcaccg ccggacatca gcgctagcgg agtgtatact ggcttactat gttggcactg atgagggtgt cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggtgcgtca gcagaatatg tgatacagga tatattccgc ttcctcgctc actgactcg tacgctcggt cgttcgactg	SEO ID: 10	DACKO-	gaactccgga	tgagcattca	tcaggcgggc	aagaatgtga	ataaaggccg
tccagctgaa cggtctggtt ataggtacat tgagcaactg actgaaatgc ctcaaaatgt tctttacgat gccattggga tatatcaacg gtggtatatc cagtgattt tttctccatt ttagcttcct tagctcctga aaatctcgat aactcaaaaa atacgcccgg tagtgatctt atttcattat ggtgaaagtt ggaacctctt acgtgccgat caacgtctca ttttcgccaa aagttggccc agggcttccc ggtatcaaca gggacaccag gatttatta ttctgcgaag tgatcttccg tcacaggtat ttattcggcg caaagtgcgt cgggtgatgc tgccaactta ctgatttagt gtatgatggt gtttttgagg tgctccagtg gcttctgttt ctatcagctg tccctcctgt tcagctactg acggggtggt gcgtaacggc aaaagcaccg ccggacatca gcgctagcgg agtgtatact ggcttactat gttggcactg atgagggtgt cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggtgcgtca gcagaatatg tgatacagga tatattccgc ttcctcgctc actgactcg tacgctcggt cgttcgactg			gatagaactt	gtgcttattt	ttctttacoo	tctttaaaaa	ggccgtaata
ctcaaaatgt tctttacgat gccattggga tatatcaacg gtggtatatc cagtgattt tttctccatt ttagcttcct tagctcctga aaatctcgat aactcaaaaa atacgcccgg tagtgatctt atttcattat ggtgaaagtt ggaacctctt acgtgccgat caacgtctca ttttcgccaa aagttggccc agggcttccc ggtatcaaca gggacaccag gatttattta ttctgcgaag tgatcttccg tcacaggtat ttattcggcg caaagtgcgt cgggtgatgc tgccaactta ctgatttagt gtatgatggt gtttttgagg tgctccagtg gcttctgtt ctatcagctg tccctcctgt tcagctactg acggggtggt gcgtaacggc aaaagcaccg ccggacatca gcgctagcgg agtgtatact ggcttactat gttggcactg atgagggtgt cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggtgcgtca gcagaatatg tgatacagga tatattccgc ttcctcgctc actgactcgc tacgctcggt cgttcgactg		WIOTIGH	tccacctca	caatctaatt	ataggtacat	tgagcaactg	actgaaatgc
cagtgattt tttctccatt ttagcttcct tagctcctga aaatctcgat aactcaaaa atacgcccgg tagtgatctt atttcattat ggtgaaagtt ggaacctctt acgtgccgat caacgtctca ttttcgccaa aagttggccc agggcttccc ggtatcaaca gggacaccag gatttattta ttctgcgaag tgatcttccg tcacaggtat ttattcggcg caaagtgcgt cgggtgatgc tgccaactta ctgatttagt gtatgatggt gtttttgagg tgctccagtg gcttctgttt ctatcagctg tccctcctgt tcagctactg acggggtggt gcgtaacggc aaaagcaccg ccggacatca gcgctagcgg agtgtatact ggcttactat gttggcactg atgagggtgt cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggtgcgtca gcagaatatg tgatacagga tatattccgc ttcctcgtc actgactcg tacgctcggt cgttcgactg			ctcaseatot	tetttacast	accattagga	tatatcaaco	gtggtatatc
aactcaaaaa atacgcccgg tagtgatctt atttcattat ggtgaaagtt ggaacctctt acgtgccgat caacgtctca ttttcgccaa aagttggccc agggcttccc ggtatcaaca gggacaccag gatttattta ttctgcgaag tgatcttccg tcacaggtat ttattcggcg caaagtgcgt cgggtgatgc tgccaactta ctgatttagt gtatgatggt gtttttgagg tgctccagtg gcttctgttt ctatcagctg tccctcctgt tcagctactg acggggtggt gcgtaacggc aaaagcaccg ccggacatca gcgctagcgg agtgtatact ggcttactat gttggcactg atgagggtgt cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggtgcgtca gcagaatatg tgatacagga tatattccgc ttcctcgctc actgactcgc tacgctcggt cgttcgactg	1		cantrattt	tttctccatt	ttagetteet	tagetectga	aaatctcgat
ggaacctctt acgtgccgat caacgtctca ttttcgccaa aagttggccc agggcttccc ggtatcaaca gggacaccag gatttattta ttctgcgaag tgatcttccg tcacaggtat ttattcggcg caaagtgcgt cgggtgatgc tgccaactta ctgatttagt gtatgatggt gtttttgagg tgctccagtg gcttctgttt ctatcagctg tccctcctgt tcagctactg acggggtggt gcgtaacggc aaaagcaccg ccggacatca gcgctagcgg agtgtatact ggcttactat gttggcactg atgagggtgt cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggtgcgtca gcagaatatg tgatacagga tatattccgc ttcctcgctc actgactcgc tacgctcggt cgttcgactg	1		aactcaaaaa	atacccccacc	tagtgatgt	atttcattat	ggtgaaagtt
agggettece ggtateaaca gggacaceag gatttattta ttetgegaag tgatettecg teacaggtat ttatteggeg caaagtgegt egggtgatge tgecaactta etgatttagt gtatgatggt gtttttgagg tgeteeagtg gettetgtt etateagetg teceteetgt teagetactg aeggggtggt gegtaacgge aaaageaceg eeggacatea gegetagegg agtgtataet ggettaetat gttggeactg atgagggtgt eagtgaagtg etteatgtgg eaggagaaaa aaggetgeae eggtgegtea geagaatatg tgatacagga tatatteege tteetegete aetgaetege taegeteggt egttegaetg	1		ggaaggtgt	acategeogy	caacutctca	ttttccccaa	aagttggccc
tgatetteeg teacaggtat ttatteggeg caaagtgegt egggtgatge tgecaactta etgatttagt gtatgatggt gtttttgagg tgetecagtg gettetgtt etateagetg teecteetgt teagetactg aeggggtggt gegtaaegge aaaageaeeg eeggacatea gegetagegg agtgtataet ggettaetat gttggeaetg atgagggtgt eagtgaagtg etteatgtgg eaggagaaaa aaggetgeae eggtgegtea geagaatatg tgataeagga tatatteege tteetegete aetgaetege taegeteggt egttegaetg			ggaacctctt	acycyccyac	addicted	catttatta	ttctgcgaag
tgccaactta ctgatttagt gtatgatggt gtttttgagg tgctccagtg gcttctgttt ctatcagctg tccctcctgt tcagctactg acggggtggt gcgtaacggc aaaagcaccg ccggacatca gcgctagcgg agtgtatact ggcttactat gttggcactg atgagggtgt cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggtgcgtca gcagaatatg tgatacagga tatattccgc ttcctcgctc actgactcgc tacgctcggt cgttcgactg	1		agggetteee	yyuaudada	ttattagga	caaactccct	caatastac
gcttctgttt ctatcagctg tccctcctgt tcagctactg acggggtggt gcgtaacggc aaaagcaccg ccggacatca gcgctagcgg agtgtatact ggcttactat gttggcactg atgagggtgt cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggtgcgtca gcagaatatg tgatacagga tatattccgc ttcctcgctc actgactcgc tacgctcggt cgttcgactg	}	İ	Lyacetteeg	ccacaggiat	atatactage cg	attttaaa	tactccaata
gcgtaacggc aaaagcaccg ccggacatca gcgctagcgg agtgtatact ggcttactat gttggcactg atgagggtgt cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggtgcgtca gcagaatatg tgatacagga tatattccgc ttcctcgctc actgactcgc tacgctcggt cgttcgactg	1		cgccaactta	ctgatttagt	grargarggt	tanactacta	accondition
ggcttactat gttggcactg atgagggtgt cagtgaagtg cttcatgtgg caggagaaaa aaggctgcac cggtgcgtca gcagaatatg tgatacagga tatattccgc ttcctcgctc actgactcgc tacgctcggt cgttcgactg	1		gcttctgttt	ctatcagctg	cccccctgt	cayctactg	acggggtggt
caggagaaaa aaggctgcac cggtgcgtca gcagaatatg tgatacagga tatattccgc ttcctcgctc actgactcgc tacgctcggt cgttcgactg	1		gcgtaacggc	aaaagcaccg	ccggacatca	gegetagegg	aytytatact
tatattccgc ttcctcgctc actgactcgc tacgctcggt cgttcgactg	ì		ggcttactat	gttggcactg	argagggtgt	cagrgaagrg	backasassas
tatattccgc ttcctcgctc actgactcgc tacgctcggt cgttcgactg	l		caggagaaaa	aaggctgcac	cggtgcgtca	gcagaatatg	tgatacagga
I I I I I I I I I I I I I I I I I I I	<u> </u>		tatattccgc	ttcctcgctc	actgactcgc	tacgctcggt	cgttcgactg
cggcgagcgg aaatggctta cgaacggggc ggagatttcc tggaagatgc			cggcgagcgg	aaatggctta	cgaacggggc	ggagatttcc	tggaagatgc

caggaagata cttaacaggg aagtgagagg gccgcggcaa agccgttttt ccataggete egececetg acaagcatea egaaatetga egeteaaate agtggtggcg aaacccgaca ggactataaa gataccaggc gtttccccct ggcggctccc tcgtgcgctc tcctgttcct gcctttcggt ttaccggtgt cattccgctg ttatggccgc gtttgtctca ttccacgcct gacactcagt tccgggtagg cagttcgctc caagctggac tgtatgcacg aaccccccgt tcagtccgac cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc cggaaagaca tgcaaaagca ccactggcag cagccactgg taattgattt agaggagtta gtcttgaagt catgcgccgg ttaaggctaa actgaaagga caagttttgg tgactgcgct cctccaagcc agttacctcg gttcaaagag ttggtagctc agagaacctt cgaaaaaccg ccctgcaagg cggttttttc gttttcagag caagagatta cgcgcagacc aaaacgatct caagaagatc atcttattaa tcagataaaa tatttctaga tttcagtgca atttatctct tcaaatgtag cacctgaagt cagccccata cgatataagt tgtaattctc atgtttgaca gcttatcatc gataagcttt aatgcggtag tttatcacag ttaaattgct aacgcagtca ggcaccgtgt atgaaatcta acaatgcgct catcgtcatc ctcggcaccg tcaccctgga tgctgtaggc ataggcttgg ttatgccggt actgccgggc ctcttgcggg atatcGGTTT CTTAGACGTC AGGTGGCACT TTtcggggaa atgtgcgcgg aacccctatt tgtttatttt tctaaataca ttcaaatatg tatccgctca tgagacaata accctgataa atgcttcaat aatattgaaa aaggaagagt atgagtattc aacatttccg tgtcgccctt attccctttt ttgcggcatt ttgccttcct gtttttgctc acccagaaac gctggtgaaa gtaaaagatg ctgaagatca gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga tccttgagag ttttcgcccc gaagaacgtt ttccaatgat gagcactttt aaagttctgc tatgtggcgc ggtattatcc cgtgttgacg ccgggcaaga gcaactcggt cgccgcatac actattctca gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg gcatgacagt aagagaatta tgcagtgctg ccataaccat gagtgataac actgcggcca acttacttct gacaacgatc ggaggaccga aggagctaac cgcttttttg cacaacatgg gggatcatgt aactcgcctt gatcgttggg aaccggagct gaatgaagcc ataccaaacg acgagcgtga caccacgatg cctTGAgcaa tggcaacaac gttgcgcaaa ctattaactg gcgaactact tactctagct tcccggcaac aattaataga ctggatggag gcggataaag ttgcaggacc acttctgcgc tcggcccttc cggctggctg gtttattgct gataaatctg gagccggtga gcgtgggtct cgcggtatca ttgcagcact ggggccagat ggtaagccct cccgtatcgt agttatctac acgacgggga gtcaggcaac tatggatgaa cgaaatagac agatcgctga gataggtgcc tcactgatta agcattggca ccaccaccac caccactaaC CCGGGACCAA GTTTACTCAT ATATACttta gattgattta aaacttcatt tttaatttaa aaggatctag gtgaagatcc tttttgataa tCTCATGACC AAAATCCCTT AACGgcatgc accattcctt gcggcggcgg tgctcaacgg cctcaaccta ctactGGGCT GCTTCCTAAT GCAGGAGTCG CATAAGGGAG AGCGTCTGGC GAAAGGGGGA TGTGCTGCAA GGCGATTAAG TTGGGTAACG CCAGGGTTTT CCCAGTCACG ACGTTGTAAA ACGACGGCCA GTGCCAAGCT TAAAAAaaat ccttagcttt cgctaaggat CTGCAGTTAT AATCTCTTTC TAATTGGCTC TAAAATCTTT ATAAGTTCTT CAGCTACAGC ATTTTTTAAA TCCATTGGAT GCAATTCCTT ATTTTTAAAT AAACTCTCTA ACTCCTCATA GCTATTAACT GTCAAATCTC CACCAAATTT TTCTGGCCTT TTTATGGTTA AAGGATATTC AAGGAAGTAT TTAGCTATCT CCATTATTGG ATTTCCTTCA ACAACTCCAG CTGGGCAGTA TGCTTTCTTT ATCTTAGCCC TAATCTCTTC TGGAGAGTCA TCAACAGCTA TAAAATTCCC TTTTGAAGAA CTCATCTTC CTTCTCCATC CAAACCCGTT AAGACAGGGT TGTGAATACA AACAACCTTT TTTGGTAAAA GCTCCCTTGC TAACATGTGT ATTTTTCTCT GCTCCATCCC TCCAACTGCA ACATCAACGC CTAAATAATG AATATCATTA ACCTGCATTA TTGGATAGAT AACTTCAGCA ACCTTTGGAT TTTCATCCTC TCTTGCTATA AGTTCCATAC TCCTTCTTGC TCTTTTTAAG GTAGTTTTTA AAGCCAATCT ATAGACATTC AGTGTATAAT CCTTATCAAG CTGGAATTCa gcgttacaag tattacacaa agttttttat gttgagaata tttttttgat ggggcgccac ttatttttga tcgttcgctc aaagAAGCGG CGCCAGGGNT

F -		GTTTTTCTTT	TCACCAGTNA	GACGGGCAAC	AGAACGCCAT	Gagcggcctc
				acagtccgca		
		ccaatagaca	cggggcatga	ctatcgtcgc	cgcacttatg	actgtcttct
		ttatcatgca	actcgtagga	caggtgccgg	cagcgcccaa	cagtcccccg
		accacadadc	ctgccaccat	acccacgccg	aaacaagcgc	cctgcaccat
		tatgttccgg	atctgcatcg	caggatgctg	ctggctaccc	tgtggaacac
ļ ļ		ctacatctct	attaacqaaq	cgctaaccgt	ttttatcagg	ctctgggagg
				tcaattatta		
_		gcaaactggc	ctcaggcatt	tgagaagcac	acggtcacac	tgcttccggt
_		agtcaataaa	ccootaaacc	agcaatagac	ataagcggct	atttaacgac
[cctaccctaa	accgacgacc	gggtcgaatt	tgctttcgaa	tttctgccat
				attcaggcgt		
}				aattacgccc		
	i			cattctgccg		
1				ccagcggcat		
	ı			aaaacggggg		
i j		accacattta	aatcaaaact	ggtgaaactc	acccagggat	taactaaaac
				accetttagg		
j		gaaaaacaca	cacatettee	gaatatatgt	gtagaaactg	ccggaaatcg
		tentantatt	cactccacac	cgatgaaaac	atttcaattt	gctcatggaa
]		aaccatataa	caacactagag	cactatccca	tatcaccacc	tcaccotctt
		}		cactateeca	caccaccago	coaccyccc
		tcattgccat	acg			
SEQ ID: 20	pACKO-	gaactccgga	tgagcattca	tcaggcgggc	aagaatgtga	ataaaggccg
`	Bla	gataaaactt	gtgcttattt	ttctttacgg	tctttaaaaa	ggccgtaata
		tccagctgaa	cggtctggtt	ataggtacat	tgagcaactg	actgaaatgc
'		ctcaaaatgt	tctttacgat	gccattggga	tatatcaacg	gtggtatatc
	ш	cagtgatttt	tttctccatt	ttagcttcct	tagctcctga	aaatctcgat
		aactcaaaaa	atacgcccgg	tagtgatctt	atttcattat	ggtgaaagtt
1		ggaacctctt	acgtgccgat	caacgtctca	ttttcgccaa	aagttggccc
		agggcttccc	ggtatcaaca	gggacaccag	gatttattta	ttctgcgaag
) 1		tgatcttccg	tcacaggtat	ttattcggcg	caaagtgcgt	cgggtgatgc
		tgccaactta	ctgatttagt	gtatgatggt	gtttttgagg	tgctccagtg
1		gcttctgttt	ctatcagctg	tccctcctgt	tcagctactg	acggggtggt
		gcgtaacggc	aaaagcaccg	ccggacatca	gcgctagcgg	agtgtatact
		ggcttactat	gttggcactg	atgagggtgt	cagtgaagtg	cttcatgtgg
ł ·		caggagaaaa	aaggctgcac	cggtgcgtca	gcagaatatg	tgatacagga
		tatattccgc	ttcctcgctc	actgactcgc	tacgctcggt	cgttcgactg
1		cggcgagcgg	aaatggctta	cgaacggggc	ggagatttcc	tggaagatgc
		caggaagata	cttaacaggg	aagtgagagg	gccgcggcaa	agccgttttt
		ccataggctc	cgccccctg	acaagcatca	cgaaatctga	cgctcaaatc
		agtggtggcg	aaacccgaca	ggactataaa	gataccaggc	gtttccccct
		gacaactccc	tcgtgcgctc	tcctgttcct	gcctttcggt	ttaccggtgt
		cattccgctg	ttatggccgc	gtttgtctca	ttccacgcct	gacactcagt
		teegggtagg	cagttcgctc	caagctggac	tgtatgcacg	aaccccccgt
		tcagtccgac	cgctgcgcct	tatccggtaa	ctatcgtctt	gagtccaacc .
Į.		cggaaagaca	tgcaaaagca	ccactggcag	cagccactgg	taattgattt
		agaggagtta	gtcttgaagt	catgcgccgg	ttaaggctaa	actgaaagga
1		caagttttgg	tgactgcgct	cctccaagcc	agttacctcg	gttcaaagag
]		ttggtagctc	agagaacctt	cgaaaaaccg	ccctgcaagg	cggttttttc
1		gttttcagag	caagagatta	cgcgcagacc	aaaacgatct	caagaagatc
		atcttattaa	tcagataaaa	tatttctaga	tttcagtgca	atttatctct
		tcaaatgtag	cacctgaagt	cagccccata	cgatataagt	tgtaattctc
		atgtttgaca	gcttatcatc	gataagcttt	aatgcggtag	tttatcacag
		ttaaattgct	aacgcagtca	ggcaccgtgt	atgaaatcta	acaatgcgct
		catcgtcatc	ctcggcaccg	tcaccctgga	tgctgtaggc	ataggcttgg
		ttatgccggt	actgccgggc	ctcttgcggg	atatcGGTTT	CTTAGACGTC
1		AGGTGGCact	tttcggggaa	atgtgcgcgg	aacccctatt	tgtttattt
		tctaaataca	ttcaaatatg	tatccgctca	tgagacaata	accctgataa

atgetteat attecett titgeggeat titgegteet titgetgeet titgetgeet titgetgeet titgetgeet titgetgeet titgetgeet titgetgeet titgetgeetge actitgetge gaggegget actitetge gaggegget actitetge gaggegget actitetge getgetatat getgetaaat citggacgg decagtge actitgetgetgetacageg gaggtagaag cecectgetg teacagagg gagtagaag accattetg getgetagge tacagagg gagtagaag cecectgetg tgagatagg gectlenter Thankourne Grancesca teagagt gagatgaag cecectgetgetgetgetgetgetgetgetgetgetgetgetget							
ACCCAGANAC ACTAGEGGAG caateggaca agactatea agactaggatg gaggeggata agattgagag accattett gettoggacat teteoggatag ctggttatt getgataaat teteoggacag catggatagg catgaggacagg catcattgat teteoggatag teatgagataggacaggataggacaggacaggacaggaca			atgcttcaat	aatattgaaa	aaggaagagt	atgagtattc	aacatttccg
ctggcgaact acttactcta gcttccggc acacattata agactggatg gagcggata aagttcaga acacattat gctcggct ttcggcttg ctggtttatt gctgataaat ctgagagcgg bagcgtgg ttcaggctgg ctggtttatt gctgataaat ctgagagcgg bagcgtggg tctcgggtgg tcastgaagg acacattgag cacattggat gaacgaata gacagatcgc tagaataggt gccacattattat attaaaact tcattgtaag tctagatgatgatgatgatgatgatgatgatgatgatgatga	ļ		tgtcgccctt	attccctttt	ttgcggcatt	ttgccttCCT	GTTTTTGCTC
ctggcgaact acttactcta gcttccggc acacattata agactggatg gagcggata aagttcaga acacattat gctcggct ttcggcttg ctggtttatt gctgataaat ctgagagcgg bagcgtgg ttcaggctgg ctggtttatt gctgataaat ctgagagcgg bagcgtggg tctcgggtgg tcastgaagg acacattgag cacattggat gaacgaata gacagatcgc tagaataggt gccacattattat attaaaact tcattgtaag tctagatgatgatgatgatgatgatgatgatgatgatgatga			ACCCAGAAAC	ACTAGtgcag	caatggcaac	aacgttgcgc	aaactattaa
ctggttatt getgataaat ctggggegg tgaggfggg tctcggggta tcattgcaga gatggatagc cataggaaat gacagatagc taacagagg gagatcagc aatatggat gaacgaata gacagatagc tgagataggt gcctcACTGA TTAAGCATTG GTAACCAGGG ACAAGTTTA CTCATATATA Ctttagattg atttaaaact tcatggtgaa gatcctttt gataatcTCA TGACCAAAAT CCCTTAAGGg catgcaccat tccttgggg ggggtgtget aacggecta acctactact GGGCTGCTC CTAATGCAGG AGTCGCATAA GGGAGAGCGG GTTTCCCAG GGGATGGC TCAAAGCGA GTAGATTGG TAAGCCAGG GTTTCCCAG TCACGACGT GTAAAACGA GGCCAGTGC AAGCTAAAA Aaaatcctta gctttcgcta aggatcTGCA GTTATAATTC CTTTCAATA GGCTATAAAA TCTTTATATAG TTCTCACTA GTTTAATTC CTTTCAATAG TGCTCAAAA ATTCTCACACA AATTTTTCG GCCTTTTTAT GGTTAAAGGA TACTCAAGG AGAATAGAT TCCTTTATT TAAATAAACT. CTCTAACTC TGCAGAAGG TACCAAAA ACGATATAGA TTCTCCATT ATTGGATTC TCTTCAACAC TCCAGAGGTGG AGAAACACA AATTTTTCG GCCTTTTTAT GGTTAAAAGGA TACCAAGGA AGAATATAAA TTCCCTTTG AAGACACAC CTTTCTTTG CAGACTAGAGA AGGATAAAA TTCCCTTTC AAGACACAC CTTTTTTTG GAAAACGCC CTTGCTAACA TGTGTATTT TACTCTCTC ACCCACACAC CCTTTCCTACAC AGGGTTGTA ATACAACAC CCTTTTTTT GAAAACGCC CTTGCTAACA TGTGTATTT TCTTTGTCTC ACCCACACAC CCTTTCCTCT TAAACGCCAACCTT TGGATTTTCA TCCTCCTCTT CATAACTTC CAGCCAACATC CTTGCTACAC TGGTTATTT TCTCTCTCTCT CTTTCAACAC CCTCCAACAC CCTTGCTACA TGGATTTTT TCTTCTCTCTC CATAACTTC CATACCCTC CTTGCTACAC TGGATTTTT TCTTCTCTCTC CATAACTTC CATACCCTCT CTTGCTCTTT TTAAGGTAGT TTTTAAACCC AATTCTAGA ATCCCTCAC ATAACCCTTA TCAAGCTGG ATCCAGAGGT CATACACTC CTTGCTCTTT TTAAGGTAGT TTTTAAACCC AATTCTAGA ACCCCTCCTTCTTCTCTCTCTCTCTCTTCTCT	i		ctggcgaact	acttactcta	gcttcccggc	aacaattaat	agactggatg
ctggttatt getgataaat ctggggegg tgaggfggg tctcggggta tcattgcaga gatggatagc cataggaaat gacagatagc taacagagg gagatcagc aatatggat gaacgaata gacagatagc tgagataggt gcctcACTGA TTAAGCATTG GTAACCAGGG ACAAGTTTA CTCATATATA Ctttagattg atttaaaact tcatggtgaa gatcctttt gataatcTCA TGACCAAAAT CCCTTAAGGg catgcaccat tccttgggg ggggtgtget aacggecta acctactact GGGCTGCTC CTAATGCAGG AGTCGCATAA GGGAGAGCGG GTTTCCCAG GGGATGGC TCAAAGCGA GTAGATTGG TAAGCCAGG GTTTCCCAG TCACGACGT GTAAAACGA GGCCAGTGC AAGCTAAAA Aaaatcctta gctttcgcta aggatcTGCA GTTATAATTC CTTTCAATA GGCTATAAAA TCTTTATATAG TTCTCACTA GTTTAATTC CTTTCAATAG TGCTCAAAA ATTCTCACACA AATTTTTCG GCCTTTTTAT GGTTAAAGGA TACTCAAGG AGAATAGAT TCCTTTATT TAAATAAACT. CTCTAACTC TGCAGAAGG TACCAAAA ACGATATAGA TTCTCCATT ATTGGATTC TCTTCAACAC TCCAGAGGTGG AGAAACACA AATTTTTCG GCCTTTTTAT GGTTAAAAGGA TACCAAGGA AGAATATAAA TTCCCTTTG AAGACACAC CTTTCTTTG CAGACTAGAGA AGGATAAAA TTCCCTTTC AAGACACAC CTTTTTTTG GAAAACGCC CTTGCTAACA TGTGTATTT TACTCTCTC ACCCACACAC CCTTTCCTACAC AGGGTTGTA ATACAACAC CCTTTTTTT GAAAACGCC CTTGCTAACA TGTGTATTT TCTTTGTCTC ACCCACACAC CCTTTCCTCT TAAACGCCAACCTT TGGATTTTCA TCCTCCTCTT CATAACTTC CAGCCAACATC CTTGCTACAC TGGTTATTT TCTCTCTCTCT CTTTCAACAC CCTCCAACAC CCTTGCTACA TGGATTTTT TCTTCTCTCTC CATAACTTC CATACCCTC CTTGCTACAC TGGATTTTT TCTTCTCTCTC CATAACTTC CATACCCTCT CTTGCTCTTT TTAAGGTAGT TTTTAAACCC AATTCTAGA ATCCCTCAC ATAACCCTTA TCAAGCTGG ATCCAGAGGT CATACACTC CTTGCTCTTT TTAAGGTAGT TTTTAAACCC AATTCTAGA ACCCCTCCTTCTTCTCTCTCTCTCTCTTCTCT	1		gaggcggata	aagttgcagg	accacttctg	cgctcggccc	ttccggctgg
teatgeage actaggagea gatetage taacagaage gagateage acatatagat gaagatege tagagataggt gectCACTGA TTAACCATTG GTAACCGGG ACCAACTTA CTCATATATA Ctttagattg atttaaaact tcattttaa tttaaaagga tctaggtag agatecgttttt gataatcTCA TGACCAAAAT CCCTTAACGG catgcaccat tcctttgggg gggtgetc aacggcctca actactact GGGCTGCTTC CTAATGCAGG AGTCGCATAA GGGGACAGCGT CTGGCGAAAG GGGATGTGC TGCAAGGCGA TTAAGTTGGG TAACGCCAGG TCACGACGTT GTAAAAGCA GGCCAGTCAA GGGGACAGCGT CTGGCGAAAG GCGATGTGC TGCAAGGCGA TTAAGTTGGG TAACGCCAGG TTATTATAAG TTCTTTATACT GGCTTATAATCT CTTTCTAATT GGCTCTAAAA TCTTTATAAG TTCTTCAGCT ACAGCATTT TTAATACAT TGCATCAAA TCCTTATATTA TTCTTCAGCT ACAGCATTT TTAATACAT TAGGACCATA ATTCCCACCA AAATTTTCCG GCCTTTTAT GGTTAAAGGA TATTCAAGGA AGTATTAAGA TATTCCATTTA AGAACTACT CTTCAACAC TCCAGCTGGG CAGTAAAAA TTCCCTTTTA AGAACTACT CTTCTTGGAC AGTCATCAAC ACCTATAAAA TTCCCTTTTA AGAACTACT CTTCTTCGCAC AGCCTGGG CCGTTAACAC AGGGTTGTGA ATACAACAA CCTTTTTTGG TAAAAGCTCC TTGGCTCTT TTAAGTGTATT TCTCTCTCCC ATCCTACAAC TCCTCCAAC ACCGTTAAAAA TATGCATTTA TCCTCTCTCT CTCTTTTTGG TAAAAGCTCC TTGGCTCTT TTAAGGTGGA ATACAAACAA CCTTTTTTGG TAAAAGCTCC TCTGGCTCTT TTAAGGTGTA TTTTAACCCT ACCTTCCACA CTCGCAACACC CTGGGACACCT TGGATTTCA TCCTCTCTTG CATTATAGCTT CATACCTCT CTTGGCTCTT TTAAGGTGTA TTTTAACCC ATCCTTCAG CATTACGTGT ATAATCCTTA TCAAGCTGGA ATTCAGCTTTAG CATTACAGGTT ATAATCCTTA TCAAGCTGGA ATTCAGCTTTTC CATTACAGGTTAG CATTACGTGT ATAATCCTTA TCAAGCTGGA ATTCAGCTTTTC CATTACAGGTTAG CATTACAGTGT ATAATCCTTA TCAAGCTGGA ATTCAGCCTATC CATTATAGC ACTTACAGTTT ATATCCTTA TCAAGCTGGA ATTCAGCCTATC CATTATAGC ACTTACAGTTT CTTGGCTCTT TTAAGGTGTA TTTTAAACCC AATCTATAGA CATTACAGTTT ATAATCCTTA TCAAGCTGGA ATTCAGCCTATC CATTATAGC ACTTACAGTTT ATAATCCTTA TCAAGCTGGA ATTCAGCTTT TCTTTTTCAC ACTTAGAGCTG GCAACAGAAC GCCATGAGGG GCCCAGGGGGGGGGGGGG			ctggtttatt	gctgataaat	ctggagccgg	tgagcgtggg	tctcgcggta
tagagatagg ggagteagge aactatggat gaacgaaata gacagatege tgagatagg gccCCTGA TTARCCGGM CACAGTTTA CTCATATATA Ctttagattg atttaaaact tcattttta tttaaaagga tctaggtgaa gatcetttt gataatcTCA TGACCAAAAT CCCTTAACGg catgaccat tccttgegg ggogtgct aacggcctca acctactact GGGCTGCTTC CTAATGCAGG AGTCGCATAA GGGAGAGCGT CTGGCGAAAG GGGGATGTTC TAATGCAGG AGTCGCATAA GGGAGAGCGT CTGGCGAAAG GGGCATTGCT GTAAAAGCAC GGCCAGTGGC AGCCTTACCAG TCACGAGGT TAAAACACAC GGCCAGTGGC AGCCTTACCAG GGCATTGTC TAAAACACAC GCCAGTGGC AGCCTTAAAA Aaaatcctta gctttcggct agggatcaca cacacacacacacacacacacacacacacacac	1		tcattgcagc	actogggcca	gatggtaagc	cctcccgtat	cgtagttatc
tgagataggt gcctCACTGA TTANGCATTG GTAACCCGGG ACCAGGTTA- CTCATATATA Ctttagattg atttaaaat tcattttaa tttaaaagga tctaggtgaa gatcetttt gataatCTCA TGACCAAAAT CCCTTAACGg GGGCTGCTTC CTAATGCAG AGTGCCATAA GGGCAGACGG TCAGAGGCG TTAAGTGGG TAAGTGGG TAAGCCAGG GTTTTCCCAG GGGATGTC TGCAAGGCGA TTAAGTTGGG TAAGCCAGG GTTTTCCCAG TCAGAGGCG TTAAGACGA GGCAGTGCA AGCGCAGG GTTTTCCCAG GGGATGTC TGCAAGGCG GTTAAGATTG GTAAAACCAT GCTTATATAAG TTCTTCAGCT ACACACTTT TTAAATCCAT TGCATCAAA TCCTTATTATAAG TTCTTCAGCT ACACACTTT TAAAATCCAT TGCATCAAA ACTCTCACAC AATTTTCCG GCCTTATAT GGTCAAAACA ACTATTATAA TTCTTTATCTT AGCCTTAAT GGTTAAAACCAT CCGTTAAGAC AATTTTCCT GCCTTATCT TCAAACCAT TAACCGCAGGG AGCATATGCTT TCTTTATCTT AGCCCTAACCAC TCTCAACAAC CCGTTAAGAC AGGGTTGTGA ATACAAACAA CCTTTTTTGG TAAAACCAC CCGTTAAGAC AGGGTTGTGA ATACAAACAA CCTTTCTTTCT TAAGACTAT CTTCACTCC ACCCCCAAA ACCCCTAAA TAAGACTAT CATTCAACCC ATCCTCCAA CTGCAACACT CAGCAACCTT TGGATTTCA TCCTCCTCC ACCCCTCCAA CTGCAACACTC CAGCAACCTT TGGATTTCA TCCTCCTCTC CATCAACACTC CAGCAACCTT TAAGACTAGA TTCCTCCTCTC CATCAACACTC CAGCAACCACACCTC TAAGACTAC CTTCTATAGCC AATCACCTCT ATAACCCTTA TAAGACTAGA TTCCTCCTCTC CATCAACACTC CAGCAACCACACACACCC CCCAAGAGG GCCCCCACAGGGG CCCAACACACAC	.		tacacgacgg	agagtcaggc	aactatggat	gaacqaaata	gacagatcgc
CTCATATATA Ctttagatega gatoctttt gataatcTCA TGACCAAAAT CCTTATAGG catgaccat tccttgegg ggcgtgctc acctactact GGGCTGCTTC CTAATGCAG AGTGCCATAA GGACAAAT CCTTAAGG GGGATTGCTC TGAAGGCAG TAAGTTGG TAAGGCAGGT TTGCCAGA GGGATTGCT GTAAAGCAC GGCCATGGC AACGTAAAA Aaaatcctta gctttcgeta aggatcTGCA GTATAATATCT CTTCTAATT GGTTCAAAA TCTTTATAAG TCCTTAATTCG GCCTTATATT TAAATAACT. TCTTATTTT TAAATAAACT. CTCTAACTCC TCATAGCTAT TAACTGCAAT ATCTCTACACA AATTTTTCG GCCTTTTTAT GGATGCAAT ACTCTAACAC AATTTTTCG GCCTTTTTTT GTAATAGCTAA ACCGTATAAAA TTCCCTTTG AAGACTCAT CTCTCAGCAGG AGTATTTAGC TACTCCATT ATTGATTTC CTCTCTGGAG AGTCATACA AGCTATAAAA TTCCCTTTG AAGACTCAT CTCTCTGGAG AGTCATCAGG AGCATTAAAA TTCCCTTTG AAGACTCAT CTCTCTGGAG AGTCACACA CCGTTAAGAC AGGTTGTGA AACAAACAA CCTTCTTGGAG AGTCACACAC CCGTAACA TAATGAATAT CATTACCTG CATTATTGGA TACAAACAT CAGCAACCTT TGGATTTTCA TCCCTCTTCTTGGAG AGTCACACAC AACGCACACTT TGGATTTTCA TCCTCTTCTTG CAACAAC CCTCACACAC AACGCACACTT TGGATTTTCA TCCTCTTCTTGGAG AGTCACACAC CAGCAACCTT TGGATTTTCA TCCTCTTCTTGGAT ATCCACACAC AACGCACACT TGGATTTTCA TCCTCTTCTTGGAT ATCCACACAC CCTGCTAACA TAAGGATGT TTTTAAACCTG CATTATTGGA TACATACTT CAGCAACCTT TGGATTTTCA TCCTCTCTTG CATTATTGGA TACATACTT CAGCAACCTT TGGATTTTCA TCCTCTCTTG CATTATTGGA TACATACTT CAGCAACCTT TAAGGTGGA ATTCAGCTG AACCTATACGTGT ATAATCCTTA TCAAGCTGA ATTCAGCTG AACCTATACGTGT CTGCTCTTT TTAAGGTGGT ATTAAACCTG CATTATTGGA TACATACTT CTTGCTCTTT TAAGGTGGA ATTCAGCTG TCTTTTCACCC AGTNAACCTG CGCACACAGAC CGCATAGGG gcctatttc tattctgag tacacacagt ccgccacacgd tatgacttt ttcagagacacagg gccaccacacacacacacacacacaca	_		tgagataggt	gcctCACTGA	TTAAGCATTG	GTAACCCGGG	ACCAAGTTTA
tcttaggtgaa gatcetttt gataatcTCA TGACCAAAAT CCCTTAAGGg catgacact tccttaggg ggoggigtca aaggcctca actateatet GGGCTGCTTC CTAATGCAG AGTGCATAA GGGAGACGCT CTGGCGAAAG GGGATGTC TGAAGGCGA TAAGTTGG TAAAGCAC GGCCAGTGC AAGCTTAAAAAAAAAA			СТСАТАТАТА	Ctttagattg	atttaaaact	tcatttttaa	tttaaaagga
catgacata tecttgagga ggagtgata aacggactaa acetactaat GGGCTCTC CTARGCAGAG GGGAGAGGGT AGGGGAAGGGT AGGGGAAGGGT AGGGGAAGGGT AGGGGAAGGGT AGGGGAAGGGT AGGGGAAGGGT AGGGGAAGGGT AGGGGAAGGGT AGGGGAAGGGT AGGGGAAGGGT AGGGAAGGGT AGGGCAAGGG GTTTTCCAGT AGAGGATTT AGAGTAAAA Aaaatactaa gctttaaaacaa agcacaa agagacaaa aacaagtta agagcacaa agagcactaa acaagatta tatagtagaa aacagtata agagcactaa agagcactaa agagcactaa acaagatta tatagtagaa agagcactaa agagcactaa acaagagta acaagagta agagcactaa agagcactaa acaagagta acaagagta agagcactaa acaagagta agagcactaa acaagagta acaagagaa agagcactaa agagcactaa acaagagta acaagaga agagcactaa agagcactaa agagcactaa acaagagta acaagagaa agagcactaa agagcactaa agagcactaa agagcactaa agagcactaa acaagagta acaagaga agagcactaa agagcactaa agagcactaa agagcactaa agagcactaa agagcactaa agagcactaa acaagagta acaagagta acaagagaga agagcacta			tctaggtgaa	gatcetttt	gataatCTCA	TGACCAAAAT	CCCTTAACGg
GGGTGCTTC CTANTGCAGG AGTCGCATAA GGGAGAGCGT CTGGCGAAAG GGGAGTGCC GTAAAACGCAG TAAAGCGCAG TCTTTCCCAG TCACGACGTT GTAAAACGAC GGCCAGTGCC AAGCTTAAAA Aaaatcctta gctttegcta aggatCTGCA GTATAAACT CTTTCATATA GGCTCTAAAA TCTTTATAAG TCTTCAGCT ACAGCATTT TAAATCCAT TGGATGCAAT ACTCCACCA AATTTTCAG TACACCAGCTTT TAAATCCAT TAGATCCAT ACTCCACCA AATTTTCTG GCCTTATAATC CTTCAACAC CAGTATGCTT TCTTATACTT AGCCCTAAT CTTCTGACACAC CAGTATGCTT TCTTATACTT AGCCCTAATC CTTCACACAC CCGTAAGAC TGGATTTCA AGCCCTAATC CTTCCACCAC CCGTAAGAC TGGATTTTA AGCACCACAC TCCACCACAC CCGTAAGAC TGGATTTTTT TCTCTGCTCC ATCCCTCCAA CTCCAACAC CCGTTAAACA TAGGATTTTT TCTCTGCTCC ATCCCTCCAA CTCCAACAC CCGTAACACTT TGGATTTTTT TCTCTGCTCC ATCCCTCCAA CTCCAACAC CAGCAACCTT TGGATTTTAT TCTCTGCTCC ATCCCTCCAA CTCCAACAC CAGCAACCTT TGGATTTTAT TCCTCTCTTC CTTAATTGG TAAAAGCTCC CTTGCACACTT TTAAGCATGAT TTTTAAACCC ACTTCAACAC CAGCAACCTT TTAAGCATGAT TTTTAAACCC ACTTCAACAC CAGCAACACA CGCAACACAC CCTTGCACACAC TCCAACACC AATAACCCTT TAAAGCATGAT TTTTAAACCC AATCTAATGA CTACAACAC CCTTGCACACAC TCCACCACAC CCCCCCAACACACA CCCCACCACACAC CCCCCCACACAC CCCCACCACACAC CCCCCCACACAC CCCCACCACCACC CCCCCACACAC CCCCACCACCACC CCCCCACACACA			catacaccat	teettacaac	gacagtactc	aacggcctca	acctactact
GGGGATCTC TGCAAGCCG TTAAACTTGG TAACCCCAG GTTTTCCAG TCACCACCTT GTAAACGAC GGCCAGTGCC AACCTTAAAA Aaaatcctta gctttcgcta aggatcTGCA GTTATAATCT CTTTCTAATT GGCTCTAAAA TCTTTATAAG TTCTTCAGCT ACACCATTTT TTAAATCCAT TGGATGCAAT TCCTTATTTT TAAAATAAACT.CTCTAACTCC TCATAGCTAGT TAACTGCAAT ACTATCCACCA AATTTTCTG GCCTTTTAT GGTTAAAGGA TATTCAAGGA AGTATTAACA TATCCCATT ATTGGATTTC CTTCATAGCAG TCAGCCTGGG CAGTATGCTT TCTTTATCTT AGCCCTAATC TCTTCTGGAG AGTCATCAAC AGCTATAAAA TCCCTTTTG AAGAACACA CTTTCTTGGA AGTCATCAAC AGCGCTAAAA TAAGGAATTT TCTTTGCTCC ATCCTCAAAC CCGTTAAGCA TGGTTATTTT TCTTTGCTCC ATCCTCAAAC ACGCCTAAA TAAGGAATAT CATTAACCTC CATTATTGGA TAAGACACC CTTGCTAACA TGTGTATTTT TCTCTTCTCC ATCCACAAC AACGCCTAAA TAAGGAATAT CATTAACCTC CATTATTGGA TAAGACACT CAGCAACCAT TGGATTTTCA TCCTCTTCT CATAAGCTC ATCCTCCAA CTGCAACATC AACGCCTAAA TAAGGAATAT TCTCTTTTC TCATAAGTTC CATTAACCTT CTTGCTCTTT TCAAGCTGGA ATTCAGCTGC ATCCTCAAA CATCACTCTTT CTTGCTCTTT TCAAGCTGGA ATTCAGCTGC ATCCTCACA CTGCAACATC AAACCCACAAGAA ACCGCGCCA ATCCACCTCAAC CTGCAACATC CCGCACCAGCA GCCACCA GGGNTGTTT TCTTTTCACC ATCACCCTTG GCAACAGAAC GCCACCA GGGNTGTTT TCTTTTCACC ATCACCCTGTG GCAACAGAAC GCCACCA GGGNTGTTT TCTTTTCACC ATCACCACTGGG GCAACAGAAC GCCACCA GGGNTGTTT TCTTTTCACC ATCACCCTGGG GCAACAGAAC GCCACCA GGGNTGTTT TCTTTTCACC ATCACCTGTG GCACACAGAAC GCCACCAC GGGNTGTTT TCTTTTCACC ATCACCCTGGG GCAACCAGAC GCCACCA GGGNTGTTT TCTTTTCACC ATCACCCACCACCACCACCACCACCACCACCACCACCACC			CCCCTCCTTC	CTAATCCACC	ACTCCCATAA	GGGAGAGCGT	CTGGCGAAAG
TCACGACGT GTAAAACGAC GGCCAGTGCC AAGCTTAAAA Aaaatctta gctttcgcta aggatcTGCa GTATAAACTC CTTTCTAATTA GCCTTAAAAA TCTCTATATAA TTCTTCAGCT ACACCATTT TTAAATCCAT TGGATGCAAT ACTCTCACCA AATTTTCTG GCCTTTATAT GTATAAGCAA ACTCACCA AATTTTCTG GCCTTTTATA GTTTAACAGAA ACTCACCACA AATTTCTCG ACCTTTTATA GTTTAACAGAA AGCAATAAAACT TCTTTATCTT AGCACTAACA TCTCACTGGG CAGTAGACA TCTTTTCTTTTATCTT AGCACTAACA CTCTACACAC CCGTTAAGAC AGGGTTGTGA ATACAAACAA CCTTTTTTGG AAGACTCCC CTTGCTAACA TGTGATTTT TCTTGGTCC ACCCCTCCAA CTGCAACAC CCGTTAAGAC AGGGTTGTGA ATACAAACAA CCTTTTTTGG TAAAGCTCC CTTGCTAACA TGTGATTTT TCTTGCTCC ACCCCCCAA CTGCAACAC CACCAACCTT TGGATTTTATCT TCTTGCTCC ATACTTCCTT CTTGCTCTT TTAAGGTAGT TCTTTCTTC TCATAAGTC CATCAACAC CACCAACCTT TGGATTTTAT TCTTGCTCC AATTATTGGA TAGATAACTT CAGCAACCTT TCAAGCTGGA ATTCAGCTG CATTATTGGA TAGATAACTT ATAATCCTTA TCAAGCTGGA ATTCAGCTG CATTATTCGA TAGATAACTT ATAATCCTTA TCAAGCTGGA ATTCAGCTGCA AATCTATGAG CATTCAGTGT ATAATCCTTA TCAAGCTGGA ATTCAGGTGT TCTTTTCACC AGTNAGCGG CCAACAGAAC GCCATCAGGG GCCCCC GGCNTCTTT TCTTTTCACC AGTNAGCGG CCAACAGAC GCCATCAGGG GCCCCC GGCNTCTTT TCTTTTCACC AGTNAGCGG GCAACAGAC GCCATCAGGG GCCCCC GGCNTCTTT TCTTTTCACC AGTNAGCGG CCGCAGAGAC GCCCATCAGGG GCCCCCC GGCCCCC GGCGCCCC GGCGCCCC GGCGCCCC GGCCACCACCCC GCCGCCCC GCCGCCCC GCCGCCCC GCCGCCCC GCCGCC			CCCCAMCMCC	TCC A A CCCC A	ተመስ አርጥጥርርር	TAACCCCACC	CTTTTTTCCCAG
getttegeta aggatchec Getatraate Cetteraate Tocttaate Tecttaate Tecttaate Acacantit Traancacta Tocttaate Acacantit Traancacta Traancacta Traancacta Traancacta Traancacta Traancacta Traancacta Traancacta Traancacta Traancacta Traancacta Traancacta Traancacta Acacanticate Getatraate Getatraate Getatraate Getatraate Getatraate Tractaate Tractaate Tractaate Tractaate Tractaate Cetatraate Tractaate Acacanticate Getatraate Cetatraate Cetatraate Cetatraate Tractaate Cetatraate			GGGGATGTGC	TGCAAGGCGA	CCCCACTCCC	AACCOCCAGG	Asstratts
TCTTTATAAG TCTTTCAGCT ACAGCATTT TTAAATCCAT TGGATGCAAT ACTCCACCA AATTTTCCTG GCTTTTTAT GGTTAAAGGA TATTCAAGGA ACTATTAGC TATCTCCATT ATTGGATTTC CTCAACAAC TCCACGTGGG CAGTATGCTT TCTTATCTT AGCCCTAATC CTTCACAAC TCCACCTAGC CAGTATAAAA TTCCCTTTG AAGAACTA CTTTCCTTC CCATCCAACA CCGTTAACAAC AGGGTTGTGA ATACAAACAA CCTTTTTTTG TAAAAGCTCC CTTGCTAACA TGGATATTT TCTTGCTCCA ATCCCTCAAC ACGCACCTC AACGCCTAAA TATGAATAT CATTAACCTG CATTATTGG TATAAGCTCC CTTGCTAACA TGGATTTTCA TCCTCTTG CATCAACAT CTGCACACTC CTTGCCTTT TAAAGGTAGT TTTTAAAGCTC CATTATTGA CATTCACTT CTTGCCTCTT TAAGGTAGT TTTTAAAGCTC CATTATTGA CATTCACTT CTTGCCTCTT TAAGGTAGT TTTTAAAGCTC AATCTATAG CATTCACTT CTTGCCTCTT TAAGGTAGT TTTAAAGCTC CATTATTGA CATTCACTT CTGCACCAGAAC GCCATGAGG GCCCCA GGGNTGTTT TCTTTTCACC AGTMAGACGG GCAACAGAAC GCCATGAGG GCCCCA GGGNTGTTT TCTTTTCACC AGTMAGACGG GCAACAGAAC GCCATGAGG GCCCCA GGGNTGTTT TCTTTTCACC AGTMAGACGG GCAACAGAAC GCCATGAGG GCCCCA GGGNTGTTT TCTTTTCACC AGTMAGACGG GCAACAGAAC GCCATGAGG GCCCCA GGGNTGTTT TCTTTTCACC AGTMAGACGG GCACCAGAC GCCCCA GGGNTGTTT TCTTTTCACC AGTMAGACGG GCACCAGAC GCCCCA GGGNTGTTT TCTTTTCACC AGTMAGACGG GCACCAGCC GCCCCA GCGNTGTTT TCTTTTCACC AGTMAGACGG GCCGCCCCC CCCGGGCCCA GCGGNTGTTT TCTTTTCACC AGTMAGACGG GCCGCCCCC CCCGGGCCCA GCGGNTGTTT TCTTTTCACC AGTMAGACGG GCCGCCCCC CCCGGGCCCA GCGGNTGTTT TCTTTTCACC AGTMAGACGG GCCGCCCCCC CCCGGGCCCA GCCCCAGCGCCCA GCCGGGCGCCACCCCCGGGCGCCCACCCCCCGGGCCCACCCCCC			TCACGACGTT	GTAAAACGAC	CUMPANA AUCU		CCCTCTAAAA
TCCTATTTT TAAATAAACT CTCTAACTCC TCATAGGTAT TAACGGTAACTCCACCA AATTTTCTG GCCTTTTTAT GGTTAAAGGA TATTGAAGGA AATTTTACT CTTCACTAT ATTGAATTC CTTCAGACAC TCCAGCTGGG CAGTATGCTAT TCTTTATCTT AGCCCTAATC CTTCAGACAC TCCAGCTGGG CAGTAAGACAC ACCTATAAAA TTCCATTTATACTT AGCCCTAATC CTTTCTGGAG AGTCATCAAC CCGTTAAAAA TCCATTTTATACTT AGCCCTAATC CTTTCTTGGAG AGTCATCAAC CCGTTAAGAC AGGGTTGTAA ATACAAACAA CCTTTTTTGG TAAAAGCTCC CTTGCTAACA TGTGTATTTT TCTCTGCTC ATCCCTCCAA CTGCAACATC CAGCAACACT TGGATTTTCA TCCTCTCTG CATTATAGGT AGAGACCTAT TTAAAGTATA CATTAAACTG CATTATAGA TAGATCACTT TTAAAGGTAGT TTTTAAAGCC AATCCATATAGA CATTCAGTGT ATAAACTCCTT CTTGCTCTTT TAAAGGTAGT TTTTAAAGCC AATCTATAGA CATTCAGTGT ATAAACTCCTT TCTTGCTCTTT TAAAGGTAGT TTTTAAAGCC AATCTATAGA CATTCAGTGT ATAAACTCCTT TCTTGCTCTTT TAAAGGTAGT TTTTAAAGCC AATCTATAGA CATTCAGTGT ATAAACTCCTT TCTTTCACC AGTNAAGACG CCAACAGAAA GCCACCAACAGAG GGCGCCA GGGNTGTTT TCTTTTCACC AGTNAAGAGG CCAACAGAAA GCCAACAGAG GGCGCCA GGGNTGTTT TCTTTTCACC AGTNAAGAGG CCAACAGAAA GCCAACAGAGAGAG	Ī		gctttcgcta	aggatCTGCA	GTTATAATCT	CITICIAATI	MCCAMCCAAM
ATCTICCACCA AATTTTCTC GCCTTTTTAT GGTTAAAGGA TATTCAGGA AGTATTAGC CAGTATGCTT TACTCCATT ACCCCTAATC TCTACACAC TCCAGCTGGG CAGTATGCAT TCTTTATACTT ACCCCTAATC TCTTCTGGAG AGTCATCAC AGCTATAAAA TTCCCTTTTG AAGAACTAC TCTTCTGGAG AGTCATCAC CCGTTAAGAC AGGTTGTGA ATACAACAA CCTTTTCGT TAAAGCTC CTTGCTAACA TGGTTATTTT TCTCTCTCC ATCCCTCCA CTGCAACACT AACGCCTAAA TAATGAATAT CATTAACCTG CATTATTGGA TAGATACTT CAGCAACCTT TGAATTTCA TCCTCCTCTT CTATAAGCTC CATTATAGA CATTATAGATACTT ATAATCCTTA TCAAGGTAGT TTTTAAACCTG CATTATAGA CATTATAGATACTT ATAATCCTTA TCAAGGTAGA ATCCAGGTG ACCACACTC AAAACACAACACACAC GCCAACAGAAC GCCATGAGG GGCTCTTT TCTTTTCACC CATCAGTGT ATAATCCTTA TCAAGGTAGA ATCCAGGTG GCAACAGAAC GCCAACAGAAC GCCATGAGG GGCTCA GGGTTTTT TCTTTTCACC AGTNAGACGG GCAACAGAAC GCCATGAGG GCCCA GGGTTTTT TCTTTTCACC AGTNAGACGG GCAACAGAAC GCCATGAGG GCCCA GGGTTTTT TCTTTTCACC AGTNAGACGG GCAACAGAAC GCCATGAGG GCCCA GGGTGTTT TCTTTTCACC AGTNAGACGG GCAACAGAAC GCCATGAGG GCCCA GGGTGTTT TCTTTTCACC AGTNAGACGG GCAACAGAAC GCCATGAGG GCCCA GGGTGTTT TCTTTTCACC AGTNAGACGG GCAACAGAAC GCCATGAGG GCCCAG GCGGTGACA GCCACAGAGAC CCCGGGACAC ACCACAGAG GCCCGGGACAC ACCACAGAG GCCCGGACAC ACCACAGAG GCCCGGGACAC ACCACAGAG ACCACAGAG ACCACACAGAC ACCACAGAG ACCACACAG ACCACAGAG ACCACACAG ACCACACAGAC ACCACAGAG ACCACACAG ACCACACACA			TCTTTATAAG	TTCTTCAGCT	ACAGCATTTT	TTAAATCCAT	TGGATGCAAT
AGTATTTAGC TATCTCCATT ATTGGATTC CTTCAACAAC TCCAGCTGGG CAGTATGATT TCTTTATCTT AGCCCTAATC TCTTCAGAGA AGCATCAGAC AGCTATAAAA TCCCCTTTTC AAGACACAC TCTTCCTACAC CCGTTAAGAC AGGGTTGGA ATACAACACA CCTTTTTGG TAAAAGCTCC CTTGCTAACA TGTGTATTTT TCTCTGCTCC ATCCCTCCAA CTGCAACACT AACGCCTAAA TAAGAATAT CATTAACCTG CATTATTGGA TAGATAACTT CAGCAACCTT TGGATTTTCA TCCTCCTTG CATTAAGTTC CATACTCCTT CTTGCTCTTT TTAAGGTAGT TTTTAAAGCTG CATTAATGATC CATACTCCTT ATAATCCTTA TCAAGCTGGA ATTCAGAGTGG ATTCAGGTGT ATAATCCTTA TCAAGCTGGA ATTCAGAGTGG ATTCAGCACAC GCAACACACAC CCATGAGG GGCATCTTTT TCTTTCACC CAGTAAGCTG ATAATCCTTA TCAAGCTGGA ATTCAGAGTGG CATACACACAC GCAACACACACAC CCCATGAGG GCCCCA GGGNTGTTT TCTTTCACC AGTNAGACGG GCAACACACACAC CCCATGAGG GCCCCATGAGG GCCACACAGC GCACACACACAC CCCCAGCAGG GCCCACAGACCC AGTNAGACGG GCAACACACAC CCCCAGCAGG GCCCACAGACCC AGTNAGACGG GCAACACACAC CCCCAGCAGC CCCGGCCCC GGGGGGGG	i i						
CAGTATGCTT TCTTTATCTT AGCCCTAATC TCTTGGAG AGTCATACA AGGTATAAAA TTCCCTTTG AAAGAACTAT CTTTCTTCT CCATCCAAAC CCGTTAAGAC AGGGTTGTGA ATACAAACAA CCTTTTTTTGG TAAAAGCTCC CTTGCTAACA TGGTATATTT TCTCTGCTCC ATCCTCCAA CTGCAACATC AACGCCTAAA TAATGAATAT TCTCTGCTCC ATCCTCCAA CTGCAACATC CAGCAACCTT TGGATTTTCA TCTCTTGTTTG CATTAATGGA TAGATACTT CAGCAACCTT TTAAGGTAGA TTTTAAACCTG CATTAATGGA ATTCACTCT CTTGCTCTTT TTAAGGTAGA TTTTAAACCTG CATTAATGGA ATTCACTCT ATAATCCTTA TCAAGCTGGA ATCAGAGGT GCAACAGAAC GCCATGAGG GGGNTGTTT TCTTTTCACC AGTMGACGG GCAACAGAAC GCCATGAGG GGCNTGTTT TCTTTTCACC AGTMGACGG GCAACAGAAC GCCATGAGG GGCNTGTTT TCTTTTCACC AGTMGACGG GCAACAGAAC GCCATGAGG GCCCA GGGNTGTTT TCTTTTCACC AGTMGACGG GCAACAGAAC GCCATGAGG GCCCATGAGG CCCGGCGCA GGGNTGTTT TCTTTTCACC AGTMAGACGG GCAACAGAAC GCCATGAGG GCCCATGAGG CCCGGGCGA CCCGGCGCA CCCGGCGCA CCCGGCGCA CCCGGCGCA CCCGGCGCA CCCGGCGCA CCCGGCGCA CCCGGCGCA CCCGGCGCA CCCGGCGCA CCCGGCGCA CCCGGCGCA CCCGGCGCA CCCGGCGCA CCCGGCGCA CCCGGCGCA CCCGGCCA CCCGGCCA CCCGGCCA CCCGGCCA CCCGGCCA CCCGGCCA CCCGGCCCCA CCCGGCCCCA CCCGGCCCA CCCGGCCCA CCCGGCCCA CCCGGCCCA CCCGGCCCA CCCGGCCCA CCCGGCCCA CCCGGCCCA CCCGGCCCA CCCGGCCCA CCCGGCCCA CCCGGCCCA CCCGGCCCA CCCGGCCCA CCCGGCCCA CCCGGCCCA CCCGGCCCA CCCGGCCCCCCCC			ATCTCCACCA	AATTTTTCTG	GCCTTTTAT	GGTTAAAGGA	TATTCAAGGA
AGCTATAAAA TTCCCTTTG AAGACTCAT CTTTCCTTC CCATCCAAC CCGTTAAGAC AGGGTTGGA ATACAACAA CCTTTTTGG TAAAAGCTC CTGCTAACA CGGGTTGAA TTACAACAA CCTTTTTGG TAAAAGCTC CAGCAACATC AACGCCTAAA TAATGAATAT CATTACCTG CATCATATGGA TAGATAACTT CAGCAACATC TTGGATTTTTA TCCTCTCTTG CATACTTTGGATTC CATCCTCTT CTTAAGTCTG TTTTAAGTAGAC AATCACTT CATACTCTT TAAGGTAGT TTTTAAAGCC AATCACTTAGA CATTCACTT ATAATCCTTA TCAAGCTGGA ATTCAGGGT CATTATAGA CATTCACTT CAGCACAGAC GGCACAGAC GGGNTGTTT TTTTTACACAGGT GACAACAGAC GCCATGAgcg gcctcattc tttgatcgt gccgcaccgt gtccggtagc tccttccggt ggggcgggg catgactatc gccgcacaca agcgcctgc accattatgt tccgaactcg ttgcggtagc cccacagtc cccggcacac accattatgt tccggatcg gccggaaca accgttttta tcaggctctg gaggcagaa taaatgaca catgactaca tctgattaa cgaacagga accattatgt tcggatcg gaatttgct tcggatcg accactcaca tctgattaa cgaacagga accattatgt tcggatcg gaatttgct tcggatcg accactcaca tctgattaa cgaacagga accattatgt tcggatcg accacacaca tctgattaa cgaacagga accactgat ccggataga cacatgat cacatgacaca tattacctca acggggta accactgat ccggataga accattacaca tcgaacaga accagcaca tagcacacaca actgacacacacacacacacacacacacacacacacacac			AGTATTTAGC	TATCTCCATT	ATTGGATTTC	CTTCAACAAC	TCCAGCTGGG
CCGTTAAGAC CCTTGCTAACA CCTTGCTAACA CCTTGCTACACA CCTGCCTAAA TAATGCATATT TCTCTCCCC AACCCCTAA TAATGCATAC CAGCAACCTT CAGCAACCTT CCTGCCTCTT TTAAGGTAGT TTTTAAACCTG ATATATTGAT CATTAACCTG ATATACCTTA TCAAGCTGAA TTTTAAACCTG ATATACCTTA TCAAGCTGAA TTTTAAAGCCC ATATATGCTC TTTAAGGTAGT TTTTAAAGCCC AATCTATAGA CATTCACTGT TTAAGGTAGT TTTTAAAGCCC AATCTATAGA CATTCACTGT TTAAGGTAGT TTTTAAAGCCC AATCTATAGA CATTCACTGT TCAAGCTGGAA TCCTCTCTT CATACCTTT TCAAGCTGGAA TTCAGGTGT TCTTTTAACC GCAACAGAAC GCCATGAgeg gcctcattt tttgatggt ccgacacgdc gtccggcagc gccgcacgcd gtccggtagc cccacacgtc gccggcaaca gcgcgcagcg cccacacgtc cccggccac gcgcgaaca agcgccttgc accattatgt tccggtacgc accattatgt tccggtacgc accattatgt tccggtacgc accattatgt tccggtacgc accattatgt tccggtacgc accattatgt tccggtacgc accattatgt tccggtacgc accattatgt tccggtacgc accattatgt tccggtacgc accattatgt tccggatccc accatcgct cccggcaca accattatgt tccggatcaca tctgtattaa cgagcctaca accatgctc ccggtacgca accattatgt tcggatacac agcacctgc gaatttgctt tcgaatttc ccggtacgca cctgacacac cctgaaccga accatgctc ggatttcac ccggtacgcac cctgaaccga accatgctc ccggtacgca cctgaaccac cctgaaccga cctgaaccac cctgaaccga accatcatc ccgagtacgc cccggccc cccgcccc cccgcccc cccgcccc cccgcccc cccgcccc cccgcccc cccgcccc cccgcccccc			CAGTATGCTT	TCTTTATCTT	AGCCCTAATC	TCTTCTGGAG	AGTCATCAAC
CTTGCTAACA TGTGTATTT TCTCTGCTCC ATCCCTCAA CTGCAACATC AACGCCTAAA TAATGAATAT CATTAACCTG CATTATTGA TAGATAACTT CAGCAACCTT TGGATTTTCA TCTCTCTTG CTATAAGTTC CATACTCCTT CTTGCTCTTT TTAAGGTAGT TTTTAAAGCC AATCATAGACCT ATAATCCTTA TCAAGGTAGT TTTTAAAGCC AATCATAGAC CATTACTGTG ATAATCCTTA TCAAGGTAGA ATTCAGGGTG AATCATAGAC CATTACTGTGT CTTGCCCTTT TTAAGGTAGT TTTTAAAGCC AATCATAGA CATTCAGTGT ATAATCCTTA TCAAGGTAGA ATTCAGGTGT AATCAGAAC GCAACAGAAC GCCATGAGG GCCCATGGTTT TCTTTCACC AGTNAGACGG GCAACAGAAC GCCATGAGG GCCCATGGTTT TCTTTTCACC AGTNAGACGG GCAACAGAAC GCCATGAGG GCCCATGGT GTTTTCACC AGTNAGACGG GCACACGAC GCCATGAGG GCCCATGGT GTTTTCACC AGTNAGACGG GCACACGAC GCCATGAGG GCCCAGTGTTT TCTTTCACC AGTNAGACGG GCACACAGCAC CCCGGCCCA GGGNTGTTT TCTTTTCACC AGTNAGACGG GCACACGAC GCCCAGACG CCCGGTCCAG GGGGGGGGGG			AGCTATAAAA	TTCCCTTTTG	AAGAACTCAT	CTTTCCTTCT	CCATCCAAAC
AACGCCTAAA CAGAACCTT CAGCAACCTT CAGCAACCTT TGGATTITCA TCCTCTCTTG CTTATAGGTTC ATTATAGTTC ATCATCATCTT ATAATCCTTA TTAAAGTAGTA ATTATAAGCC AATTATAGCA ATTATACCTT ATAAATCCTTA TTAAAGTAGTA ATTATAAGCA ATTATACCTT ATAAATCCTTA TTAAAGTAGTA ATTATAAGCA ATTATAAGCA ATTATACCTTA TTAAAGTAGAATA TAATCATTA TCAAAGTGGA ATTCAGGTGT ATTATAGACCA ATTATAACCA GCAACAGAAC GCCATCAGAG GCAACAGAAC GCCATCAGAG GCAACAGAAC GCCATCAGAG GCAACAGAAC GCCATCAGAG GCAACAGAAC GCCATCAGAG GCAACAGAC GCCATCAGAG GCAACAGAC GCCATCAGAG GCAACAGAC GCCATCAGAG GCAACAGAC GCCATCAGAG GCACACAGAC GCCATCAGAG GCACACAGAC GCCATCAGAG GCACACAGAC GCCATCAGAG GCACACAGAC GCCATCAGAG GCACCATCAC ATCACCACCAG ATTATACCTC CTATAACCTC CTATAAACCC CATTATAGCT CAATCACCTC CATTATAGCT CAATCACCAC AACAAAACCTC CATTACACCCAC AACACACCAC AACAAAACCTC TTCGCGCGAT CACCACCACA CACCACCACA CACCACCACA CACCAC			CCGTTAAGAC	AGGGTTGTGA	ATACAAACAA	CCTTTTTTGG	TAAAAGCTCC
CAGCAACCTT CTTGCTCTTT CTTGCTCTTT TTAAGCTAGT TTTTAAAGCC AATCTATAGAC AATATCATAAGTGT TAAACCCTA AATATCCTTA TAAACCCGA ATTCAGGGA ATTCAGGGA ATTCAGGGA ATTCAGGGA ATTCAGGGA ATTCAGGGA ATTCAGGGA ATTCAGGGA ATTCAGGGA CGCACAGGAC GCAACAGAAC GCCATGAGGG GCAACAGGAC GCCATGAGGG GCAACAGGAC GCCATGAGGG GCAACAGGAC GCCATGAGGG GCAACAGGAC GCCATGAGGG GCAACAGGAC GCCATGAGGG GCAACAGGAC GCCATGAGGG GCAACAGGAC GCCATGAGGG GCAACAGGAC GCCATGAGGG GCAACAGGAC GCCATGAGGG GCAACAGGAC GCCAGGGGCCCA GGGGTGTTT TCTTTTCACC ATTATCACC ATTATCACCAC ATTATCACCGG GCAACAGGAC GCCATGAGGG GCCATGCGG GCCATGCGG GCCATGCGG GCCATGCGG ACACAGAACTC ATCATCAGAAG AGGATCCCA ACCAAAAACTC ATCACGAGAG TTGGGGCCGA CCAAAAACCT TTGGGGGCCGA CCAAAAACCT TTGGGGGCCGA CCAAAAACCT TTGGGGGCCGA CCAAAAACCT TTGGGGGCCGA CCAAAAACCT TTGGGGGCCGA CCAAAAACCT TTGGGGGCCGA CCAAAAACCT TTGGGGCCGA CCAAAAACCT TTGGGGGCCGA CCAAAAACCT TTGGGGGCCGA CCAAAAACCT TTTGGCGGGAT CCGAACAGGG CCCCATGCCG ACCACCAGAG TTCCCACCTGA CCCCATGCCG ACCACCAGAG TTCCCACCTGC CCCCATGCCG ACCACCAGAG TTCCCACCTGA CCCCATGCCG ACCACCAGAG TTCCCACCTGA CCCCATCCCG ACCACCAGAG TTCACCCGA TTCACCCGA CCCCATCCCG ACCACCAGAG TTCCCACCTGC CCCACCCGA CCCCATCCCG ACCACCCGA CCCCATCCCC ACCCCATCCCG ACCCCACCAGAGAGG TTCCCACCCGA CCCCATCCCC ACCCCATCCCG ACCCCACCAGAGAGG TTCCCACCCGA CCCCATCCCC ACCCCACCCGACC ACCCCATCCCC ACCCCACCACAC ACCCCACCCGACC ACCACCCAC			CTTGCTAACA	TGTGTATTTT	TCTCTGCTCC	ATCCCTCCAA	CTGCAACATC
CAGCAACCTT CTTGCTCTTT CTTGCTCTTT TTAAGCTAGT TTTTAAAGCC AATCTATAGAC AATATCATAAGTGT TAAACCCTA AATATCCTTA TAAACCCGA ATTCAGGGA ATTCAGGGA ATTCAGGGA ATTCAGGGA ATTCAGGGA ATTCAGGGA ATTCAGGGA ATTCAGGGA ATTCAGGGA CGCACAGGAC GCAACAGAAC GCCATGAGGG GCAACAGGAC GCCATGAGGG GCAACAGGAC GCCATGAGGG GCAACAGGAC GCCATGAGGG GCAACAGGAC GCCATGAGGG GCAACAGGAC GCCATGAGGG GCAACAGGAC GCCATGAGGG GCAACAGGAC GCCATGAGGG GCAACAGGAC GCCATGAGGG GCAACAGGAC GCCATGAGGG GCAACAGGAC GCCAGGGGCCCA GGGGTGTTT TCTTTTCACC ATTATCACC ATTATCACCAC ATTATCACCGG GCAACAGGAC GCCATGAGGG GCCATGCGG GCCATGCGG GCCATGCGG GCCATGCGG ACACAGAACTC ATCATCAGAAG AGGATCCCA ACCAAAAACTC ATCACGAGAG TTGGGGCCGA CCAAAAACCT TTGGGGGCCGA CCAAAAACCT TTGGGGGCCGA CCAAAAACCT TTGGGGGCCGA CCAAAAACCT TTGGGGGCCGA CCAAAAACCT TTGGGGGCCGA CCAAAAACCT TTGGGGGCCGA CCAAAAACCT TTGGGGCCGA CCAAAAACCT TTGGGGGCCGA CCAAAAACCT TTGGGGGCCGA CCAAAAACCT TTTGGCGGGAT CCGAACAGGG CCCCATGCCG ACCACCAGAG TTCCCACCTGA CCCCATGCCG ACCACCAGAG TTCCCACCTGC CCCCATGCCG ACCACCAGAG TTCCCACCTGA CCCCATGCCG ACCACCAGAG TTCCCACCTGA CCCCATCCCG ACCACCAGAG TTCACCCGA TTCACCCGA CCCCATCCCG ACCACCAGAG TTCCCACCTGC CCCACCCGA CCCCATCCCG ACCACCCGA CCCCATCCCC ACCCCATCCCG ACCCCACCAGAGAGG TTCCCACCCGA CCCCATCCCC ACCCCATCCCG ACCCCACCAGAGAGG TTCCCACCCGA CCCCATCCCC ACCCCACCCGACC ACCCCATCCCC ACCCCACCACAC ACCCCACCCGACC ACCACCCAC			AACGCCTAAA	TAATGAATAT	CATTAACCTG	CATTATTGGA	TAGATAACTT
CTTGCTCTTT TTAAGGTAGT TTTTAAGGCC AATCTATAGA CATTCAGTGT ATAATCCTTA TCAAGCTGGA ATTCAGGGTT CACAGCTGGA ATTCAGGGTT tttatgttga cgctcaaaga AGCGGCCCA GGGNTGTTT TCTTTTCACC AGTNAGACGG GCAACAGAAC GCCATGAgeg gcctcattct ccgcaccgct gtccggtagc ccctctcggt ggggcggggg catgactact gtcgcgcac ttatgactgt cttcttatc atgacactgt tcgcgcacacacacacacacacacacacacacacacacac	•		CAGCAACCTT	TGGATTTTCA	TCCTCTCTTG	CTATAAGTTC	CATACTCCTT
ATAATCCTTA TCAAGCTGGA ATTCagogtt acaagtatta cacaaagttt tttatgttga gaatattttt ttgatgggg gocactattt tttgatcgtt cgctcaaaga AGCGGCGCCA GGGNTGTTTT TCTTTTACAC AGTNAGACGG GCAACAGAAC GCCATGAgcg goctcattct ttatctgag ttacaacagt gtcgccgcac gtcggaagc cccaacagtc ccctgggag ggggggggg catgactatc gccggaaaca agcgcctga accattatgt tcggacagg tgctgcggaaca accattatgt tcggacagg tgctgcggaaca accattatgt tcaggacagg cacatacca accattatat tatacctcc agcggaaga cacattatgt tcggatcga accattatgt tcggatcga accattatgt tatacctca accatgggaaga cctgagcaac accatgct ccggtagtca accattgt tcgaattac ggagacagaa tataccacaa tataccacaa accatgct tcggaacaacaga cacattgt tcgaattata cgaacacaga cacattgt tcgaattata acgggaaga cacattatgt tatacctca accatgct ccggtagtca acaactgct cacacactgct cacacacgga accacacaga cacactgct tcgaattata cgaaccaga cacactgct tcgaattata acaaccagga accacacaca tatacaccaa gggatatta aggacacaa taaaccggt accacacaa tacaccaca acggcacaacaa accaccaca cgcacacaca			CTTGCTCTTT	TTAAGGTAGT	TTTTAAAGCC	AATCTATAGA	CATTCAGTGT
tttatgttga gaatatttt ttgatgggc gccactatt tttgatcgtt cgctcaaaga AGCGGCGCCA GGGNTGTTT TCTTTTCACC AGTNAGACGG GCAACAGAAC GCCATGAgcg gcctcattc ttattctgag ttacaacagt cgcgcaccgct gtccggtagc tccttctggt gggcgcggg catgactatc gtcgcgcacc gccgcaccacagtc cccacacagtc cccacacagtc cccacacagtc cccacacagtc ttatacacca agcgccctgc accattatgt tccggatcg catgactatc tattacctcc acceptttta tcaggctctg gaggcagaa tcattcatattattacctcc acceptttta acgacactaca tctgtattaa cgaaccggt tagacacaggt accactacat tattacctcc acceptage accattattat tctgacagg catgacatattagacataag gaatttgctt ccggtagtaa acacacgac cccacacagtc cccggcaccg gggcctgag catacacaattatattacatcc aggcgtatta acgacacaga ctggcctcag gaatttgaga aacacacagg tagacacaagg cacactactat tctgtattaa cgaaccgaa cggcgtatta aggacacaga ctggcctcag gaatttgct ggcattata acgacccgc ccgcccgccc tgccactcat gccattcatc cgcttattat aggcgatagac cgcccggcc tgccactcat cgcagtactg ttgtaattca ttaagcattc tgccgacatag ggcatcaca gaggcgtta agggcacaa taactgcctt aaaaaaaatta taggcaaga gagcacaca gaggggcgaag acctacaca taactgcctt aaaacacaca gggggcgaag acctacaca ttgcggataaa tatttgccaa taaggcatca taaggcactca taaggcacta ttaaggaacct gggagcagaa acctacacca gggggcgaag aactcaccca aggcgttaaa tatttgccaa tagggaaaa acctacacca aggcggttggaaaacctaa tattggcaa tattggcaac tattggcaac tattggcaac tattggcaac tattgggaaaa aactcaccca aggcattggct gagaagaaaa acattatcc aaaaccgtt tcaccgtaaa aactgcgga aactgcggg ttaacacagg ttaacacagg tgaacacata cttgcgaata tattggcac agttgcca tattggcac ggttaacacct cagacgcact caggcacata gattgccgg gaacacaca ccagccacat cttgcgaata acctaccacca aggccacaccacca ggctttcat ggaaaaccgg gtgaacacata cagacgcatg gattgcca accaccacca aggccacac ccagacgcaccaccacca aggccaccaccaccaccaccaccaccaccaccaccaccac			ATAATCCTTA	TCAAGCTGGA	ATTCagcgtt	acaagtatta	cacaaagttt
cgctcaaagA AGCGGCGCA GGGNTGTTT TCTTTCACC AGTNAGACGG GCAACAGAAC CCCATCAGgcg gcctcattc ttattctgag ttacaacagt gtcgccgcac gtccgtage tccttccggt ggggcggggg catgactatc gtcgcgcaaca ttatgactgt ctctttatc atgcaactcg taggacaggt cccaacagtc cccaggcac ggggcctgcc accataccca agcgcgtagac taccetgtgg acacactaca tctgtatta cacggggt tactcctttat tcaggactag accgtttta tcaggctctg ggggcagaa tctgcagggagg taccactagt tctgtattaa cgaagcgcta accggggagag catgactaca tctgtattaa cgaagcgcta accggggagag cacactacatacca aggggagag cacactacatacca aggggagag cacactacatacca aggggagagag cctgagcaa taaatgatca tatggcaatacaggactacacaggt tagacacaggat cacactgctt cgggtagtaa ataaaccggt aaaccagaaa cgggcattata aggacacaa cacactgact cggctattata ggcggtagcac cggcattta aggacacaa taaatgaccag cgcctggactaggaattattct gcaattcat ggcgtattata ggcaggtta aggagcacaa taaatgacta tatagcatta tggcaatca ggagcactaa ggagcacaa taggcaggt ttggaaaacaggggggggaga aacctaacca aggtgtca tattggcaca gatgaaccta tagggaata aggtgtcca tattggcaca tatttgccaa tattggcaca aacctgacaca taggggggaaa aactcaccaa aggcaggtt ttcaccgtaa aacactgacaga cggggtggaaa aactcaccaa aggcaggtt ttcaccgtaa cacacgcacat cttgggaata tattggcac gattggct gaaaacaggt ttcaccgtaa cacacgcacat cttgcgaata tattggtgaa aactcaccaa aggcaggtt ttcaccgtaa cacacgcacat cttgcgaata tattggtgaa aactgcggg tggaacacgggt ttcaccgtaa ggaacacagg tggaacacata cacacgcacat cttgcgaata tattggtgaa aactgccgga aactgcgggg tggaacacata cacacgcacat cttgcgaata tattggtgaa aactgccgga ggttggaaacgggggggggg			trtatottga	gaatatttt	ttgatgggg	gccacttatt	tttgatcgtt
GCAACAGAAC ccgcaccgct gtccggtagc tccttccggt gggcgcgggg gtcgcaccgc ttatgactgt ctctttatc gggcgcgggg gccggaacacagc cccaacagtc ccccggcac ggggcctgc accatacca cgccgaaaca agcgcctgc ccccggcac ggggcctgc accatacca tgctgctggc taccctgtgg accatatagt tccggatctg accgtttta tcaggctctg ggggcagaa tcttgtattaa cgaagcgta tattacctcc acgggagag cctggcaac ctggcacag agcacacggt cacactgct ccggtagtaa accatacaa tattacctcc acgggagag cctggcaaca ctggaccaga agcacacggt cacactgctt cgggtagtaa ataaaccggt acacactgact tagacataag cggctatta agggcaccaa ctggaccaga agcgtagcac caggcgttta agggcaccaa ccggaaccgggtc gaatttgctt tcgaatttct ggcattcatc ggcgtagcac caggcgtta agggcaccaa cactgcctt aggcgacatag ggagcactaa cggacctgc cctgaaccga cgaccgggtc ggcatcagca cggcattta agggcaccaa cactgcctt tagacatata cgccccgccc tgccactcat cgcagtactg ttgtaattca tatagcata tgcgcacatg ggagcacata cagacggat gatgaacctg ggcatcagaa agggcactaa cagacggat gatgaacctg ggggtagaaa agggcacataa tattgccaa tattgccaa tattgccaa aactaccac ggggggagaa aggtgtcca tattggcac gatgaacctg gatgaacctg gggggaaa aggcaaggt ttcaccataca tattgccaa tattgccaa tattgccaa tattgccaa tattgccaa tattgccaa aactaccca ggggggaaa aactgccgga aggacgaaaa acatattcc cactattata acctaccca gggattggct gagacacaa cactactcc tataggcaaca cactacctca tattgccaa tattgccaa tattgccaa tattgccaa aactaccca gggatgaaaa acatattcca tattgccaa tattgccaa tattgccaa tattgccaa tattgccaa tattgccaa tattgccaa aactaccca ggatgaaaa acatattcca cacgcacact ttcaccgtaa aactaccct tattgcgaaaa cactaccca ggatgaacaccaa cttgcgaata aactaccca ggatgaacaccaa cttgcgaata aactaccca ggatgaacacaa cacgcacaa cttgcgaaaaa cactaccca ggatgaacacaa cacgcacaa cttgcgaaaaa aactaccca ggagcgaga gagacacaa aactaccca ggagcgagaa aactaccca ggagcacaa aactaccca ggagcacaa aactaccca ggagcacaa aactaccca ggagcacaa aactaccca ggagcacaa aactaccca ggagcacaa aactaccca ggagcacaaa aactaccca ggagcacaa aactaccca ggagcacaa aactaccca ggagcacaaa aactaccca ggagcacaaa aactaccca gagacacaa aactaccca gagacacaa aactaccca gagacacaa aactaccca gagacacaa aactaccca gagacacaa aactaccca gagacacaa aactaccaa aactaccaa aactaccaa aactaccaa aact			cactcaaagA	AGCGGCGCCA	GGGNTGTTTT	TCTTTTCACC	AGTNAGACGG
ccgcaccgct gtccgtagc tccttccggt gggcgcggg catgactatc gtcgccgcac ttatgactgt cttctttate atgcaactcg taggacaggt gccggagac cccaacagtc ccccgcacac ggggcctgc accataccaca agcgcctgc accattatgt tccggatctg catcgcagga accgttttata tcaggctctg accattatgt tccggatctg catcgcagga accgttttata tcaggctctg accatcaca tctgtattaa cgaaccga agcgctattatatatatatctcc acggggagaga cctggagaga accatcaca tctgtattaa cgaacggtatatatatatatctcc acggggagaga cctggagaaa taaatagatca tatcgcaat agcacacggt cacactgct ccggtagtca ataaaccggt aaaccaggacaacaggt tcggatatta acgacctgc cctgaaccga cctgaaccga agcacagggtt aggggagaacaca cctgaaccga cctgaaccga cctgaaccga cctgaaccga cctgaaccga cgaccgggtc aaaaccagcaa cggctattaa aggcacacaa tatgcctat aggcgcacaca aggcgcacaa acatgcct aggcaccacac ggcatcagca cgagcgctatatatggcacacacacggggggggagaagaacacacac			GCAACAGAAC	GCCATGAgcg	gcctcatttc	ttattctgag	ttacaacagt
gteggeageg cecaacagte ceceggeae ggggeetgee accatacea geggeetge accatacea agegeetge accatatgt teeggatetg categoagga accattatgt teeggatetg categoagga accattatgt teeggatetg categoagga accattatgt teeggatetg categoagga accattatgt teeggatetg categoagga accattatgt teeggatetg categoagga accattatgt teeggatetg gaagegeta taaatgatea tategoagga accattget teegagetetg ggaggeagaa taaatgatea tategoaga agoogacacaggat teegaattet teegaattet gecatteate ggegtageae eaggegotta aggacetge eeggatetg gaattget teegaattet ggeatteate eggetatatat eageoggte eaggegotta aggegoaga gageacaa eaggegotta aggegoaga gageacaa eaggegotta aggegoaga ggaggeaga eetggeateg ggagacaa taaatgacea eagaceggat gecatteate eggetatatat eagtgetge eagaceage eggecoaga eaggegotta aggeacaaa taactgeett tegaattet eggagaceaa eagacegeat gagaacetg ggaggegaaga eetggeated eagaceggat tegeaatea tattgeea gggggggaaa aactacacea ggggtteea aaggegaaaa acatattee aaaacegtea eagacegaaa acatatee eagacegaa aactacacea gggattgee gagacaaaaa acatatee eagacegat tecaacea eagacegatg tecaacegeacaa eactacee eagacegaaa aactacee eagacegaaa aactacee eagacegaaa aactacee eagacegaaa acatatee eagacegaa aactacee eagacegaaa acatatee eagacegaa aactacee eagacegaa aactacee eagacegaa aactacee eagacegaa eagaceaaa eactacee eagacegaa eagaceaaa eactacee eagacegaa eagaceaaa eactacee eagacegaa eactacee eagacegaa eactacee eagacegaa eactacee eagacegaa eactacee eagaceaaceace eagaceaace eagaceaace eagaceaaceaceaceace eagaceaaceaceaceaceaceaceaceaceaceaceaceace			ccacaccact	atccaataac	teetteegat	aaacacaaaa	catgactatc
gccgcaaca agccctgc accatatgt tccgatcta accatacca tgctgctgcg accatatgt tccgatcta tcatgcagga accatatatgt tccgatcta accatgat tctgtatta accgttttta tattacetcc agcgctatta acgacctgc cacactgct tcgagcaaa ctggcctag gcatttgaga agcacacgg cacactgct tcgagcaaa ctggcctag gcatttgaga agcacacgg cacactgct tcgagcaaa ctggcctcag gcatttgaga acgacctgc cacactgct tcgagtagaa ccggaacgaa ccggcagaa acacacagga cacactgct tcgagcacaa ctggcctcag gcatttgaga acgacctgc cgctattatat cacttattca aggcgtata agggcaccaa cggccgggtc cgcattcatc cgccocgccc tgccactcat cgcagtatta tatgcatata tatgcaatta tgccactcat cgcagtatta tatgcacat cggagacaca aggggtta agggagacaca acatgcct ttgcaattat tatgcaatta tatgcacat cgcaggat ttgcaatca tatgccag ggaggggaag aagttgcca agggggcgaag aagttgcca tatgggaacat tattgcca tggggagaacaca cattgccta tggggaacac agggggggaag aagttgcca agggattggct tatgggaaaa acatattctc aataaaccct ttagggaaat tatggtaaa acatattctc aataaaccct ttagggaaaa acatattctc aataaaccct ttagggaaaa acatattctc aataaccct tatgggaaaa acatattctc aataaaccct tatgggaaacggatg gtgaacacga gggattggc gaacggaata acatattctc aataaaccct tatgggaaacggatg gtgaacacgg gtgaacacata cacgcacat cttgcgaata acatactcac cagagcgatg gtgaacacta ggtatcaca cacgcacat ttcgcaata acatactcca cagagcgatg gtgaacacta cacgcacat ttcgcaata acatactcca cagagcgatg gtgaacacta cacgcacat ttcgcaatac cacgcacat ttcgcaatac cacgcacat ttcgcaatac acacgcacat ttcacggaatacacca acacgcacat ttcacggaatacacca acacgcacat ttcacgaacaccacacacacacacacacacacacacacac			atcaccacac	ttatgactgt	cttctttatc	atocaactco	taggacaggt
cgccgaaaca tgcgcctgc accattatgt tccggatctg catcgcagga tgctgtgtgc taccctgtgg aacactaca tctgtattaa cgaagcgcta taaatgatca tattgctaat tattacetce agcactgct cacactgctt caggaggaag cctgagcaaa ctggcctaag gaatttgctt ggcgtattaa aggacactgc cggtagtca ataaaccggt aaaccaggat gaatttgctt ggcgtatta ggcattcatc ggcgtatta ggcgtatta caggcgtta taaatgaccagga aacacaggat cacactgct cggtagtca ataaaccggt aaaccagcaa aggcgctat aggcgcacaa taggcgcta taaatgccgt gaatttgaga aaggcgacaa ccggaggtc cacactgat cggcattata cacttattca aggcgtatta ggcatcaac cgcccgccc tgccactcat cgcagtactg ggcatcaac agggggggaag aagttgcca gggggggaag aagttgcca ttgggaaaca gggggggaag aactcaccaa gggggtgaaa aactcaccca ttagggaaac tattggccaa tatttggccaa aatcgccagg gggggggaag aagttgcca ttgggaaaaca acatattccaaacacagggggggaag aactcacgaggat ttcaccgaacacacactttgggaaaacgggaaaaacggggggaagaaaacggggaaagaaacgggaaaaacggggaaaaacgggaaaaacggggaaaaacgggaaaaacgggaaaaacggggaaaaacgggaaaacgaaacacacacacacaaca			accaccacac	cccaacactc	ccccaaccac	aggacctacc	accataccca
tgetgetgge accettetgg aacacetaca tetgtattaa cgaagegeta teaggetetg ggaggeagaa taaatgatea tategteaat tattacetee aggggagag cetgageaaa etaggeeteag geatttgaga aggeetatta aggacetge cetgageaaa etagacegaa etagacataag ggatttgata acgacetge cetgageaaa etagacegaa etagacetaag gaatttgett tegaatttet geeatteate eggeetattaa eggegetatta eggegetatta eggegeteaggget etageegacatg gaageeatea eggegeteta eggeateaggeatg gattageet taaategeeta eagacegget etageegacatg gaageeatea eggegeteaggeatg gatgaacetg eggeateegggegggggggggggggaag aagttgeea tattggeeaaa eaactggeeta taagggagaaaa eaactgeeta taaaceget etageegacatg gatgaacetg gatgaacetg gatgaacetg eggeateeggggggggggggggaag aggttggee eggaggggaaaaaaaaaa			caccassacs	accectec	accattatat	tccggatctg	catcgcagga
accgtttta tcaggctctg ggaggcagaa taaatgatca tattgctaat tattacctcc acggggagag cctgagcaaa ctggcctcag gcatttgaga acgacacggt cacactgctt ccggtagtca ataaaccggt aaaccagcaa cggctattta acgaccctgc cctgaaccag gaatttgctt tcgaatttct ggcgtagtca acgacctgc cctgaaccga cgaccgggtc cacactgatt cgcattcatc cgcttattat cacttattca ggcgtagcac caggcgttta agggcaccaa taactgcctt aaaaaaaatta tgcgcgacatg ggcatcagca cgcagtcat cgcagtactg ttgtaattca ttaagcattc tgccactcat cgcagtactg ttgtaatca ttaagcattc ggaggggaag aactcacca ggggttgcc cagacagca ttgtgaatac tattgccaa tattgccaa tattgccaa tattgccaa tattgccaa tattggcaagca aactcaccca ttagggaaaa aggcagata tattggccac ggttaaaacctg gggattggct gagacgaaaa acatattctc aacaccct ttagggaaat tattgtgtaga aactgccgga aactgccgga aactgccgga aactgccgga aactgccgga aactgccgga aactgccgga aactgccgga aactgccgga aactgccgga aactgccgga aactgccgga aggacgaacg ggaggagggt ttcaccgtaa cacgccacat cttgcgaata tattgtgtaa aactgccgga aactgccgga ggaacaacgg tgtaacaagg gtgaacacta ccaggccacat tccccatatca ccagccaca ttggaaaacgg tgtaacaagg gtgaacacta ggctttcatt ggcatacg SEQ ID: 21 pKQ ATGGATCCGA GCTCGAGATC TGCAGCTGGT ACCATATGGG AATTCGAAGC GACCATCATC ATCATCATCA TTGAGGTTTAA ACGGTCTCCA GCTTGGCGGTG TTTGGCGGGAT CTGAAAAACTC TTGAGGTTTAA ACGGTCTCCA GCTTGGCGGTG TTCCAGCAGG CCGATAAAAAC AGAATTTGC TGGCGGCAGT AGCGCCGTG TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT AGCGCCGGTGG TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT			tactactac	taccetataa	accaccacge	tctgtattaa	cgaagcgcta
tattacetec aggggagag cetgageaa etggeeteag geatttgaga ageacaeggt cacactgett ceggtagtea ataaaceggt aaaceageaa eeggetatta aegaceetge cetgaacega eegtatteet geatteete geatteate eggetateat eegeateete eggeageaete eggeageaete eggeageaete eggeageaete eggeageaete eggeageaete eggeageaete eggeageaete eggeageaete eggeageaete eggeageaete eggeageaete eggeageaete eggeageaete eggeageaete eggeggeaetea eggegggeaete eggeggggggageaeae eggeggggggagagaaa eacteaeeae egggggggagagaaa eacteaeeae eggggtteetateete eggeggaaaee egggggggagagaaae eactateee eggggaaaae eactateee eggggaaaae eactateee eacegeaeae eacegeaeae eacegeage eggggggagaaaa eacatateee eacegeaeae eaceaeae eacegeaeae eacegeaeae eacegeaeae eaceaeae eacegeaeae eacetaeeae eggeggaaeaeae eacegeaeae eacetaeeae eggeggaaeaeaeaeaeaeaeaeaeaeaeaeaea			cyclyclyc	taccectyty	aacacccaca	taaatratra	tatcotcaat
agcacacggt cacactgctt ccggtagtca ataaaccggt aaaccagcaa cggctattta acgacctgc cctgaaccga cgaccgggtc gaatttgctt tcgaatttct gccattcatc cgcttattat cacttattca agcgcacaa aggcgacacaa taactgcctt aaaaaaatta taactgcgcacaa tgcgcacaca tgcgcacaca tgcgcacacacacacacacacacacacacacacacacaca			accgtttta	ceaggetety	ggaggcagaa	chaacgacca	gcatttgaga
tagacataag cggctattta acgaccetge cetgaacega cgaccgggte gaatttgett tegaattet gecatteate cgcttattat cacttatea aggegtageae caggegtta agggcaceaa taactgeett taageatte tgecgacatg gaageatea cagacggeat tgecacteat tgecgacatg gatgaacetg gatgaacetg gagggggaag aggttgee ttgegtataa tatttgeea tggtgaaaae aggggggaag aagttgee tattggeae gagacgaaa acatatete tatgggaaat tatgggaaa tattggeae ttaagggaat tatgggaaat tatggtgaa aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga gtgaacacata caggegatg tgaacaagg gtgaacacta cagacgatg tggaacacgg tgtaacaagg gtgaacacta teccatatea ceageteae gtgtteeat ggecataeg stgaacaagg gtgaacacta cagacgeatg tggaacacgg tgtaacaagg gtgaacacta teccatatea ceageteaee gtetteatt gecataeg stgaacacgg tgtaacaagg gtgaacacta teccatatea cagaccgatg tgtaacaagg gtgaacacta teccatatea teccatatea tecageteaee gtetteatt gecataeg stgaacacga tgtaacacgg tgtaacacgg tgtaacacgg tgtaacacgg tgtaacacgg tgtaacacgg tgtaacacgg tgtaacacta teccatatea teccatata t			tattacctcc	acggggagag	cctgagcaaa	ctggccccag	gcaccegaga
gaatttgett ggegtagee caggegttta agggeaceaa taactgeett taageatte tgeegacatg gaageeteat egeagtetg ttgeaatee ttgeegaetg ggegtageae aggtgeet ttgegtataa tatttgeeea ggggggegaag aagttgee tattggeeae tattggeeae tattggeae tattggeaee tattggeaae aacteaeee ggggattgget gagaegaaaa eatattee ttgggaaaee tattggeeae tattggeae tattggeae aactggtga aacteaeeee ttgggaaaae tattggeeae tattggeae aactggtga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aactgeegga aggttgee tggaaaaegg gtgaacaeta ggaaaeaegg tegaaeaeggegatg tggaaaaeeggegatg tggaaeaeeg geeataeggegatg tggaaeaeeg geeataeggegatg tggaaeaeeg geeataeggegatg tggaaeaeeg geeataeggegatg tggaaeaeeta geeataeggegatg tggaaeaeeg geeataeggegatg tggaaeaeeg geeataeggegatg tggaaeaeeg geeataeggegatg tggaaeaeeta geeataeggegatg tggaaeaeeta geeataeggegatg tggaaeaeeta geeataeggegatg tggaaeaeeta geeataeggegatg tggaaeaeeta geeataeggegatg geeataeggegatg geeataeggegatggaaeaeggegatg tggaaeaeeta geeataeggegatggaaeaeggegatggaaeaeaeggegatggaaeaeaeggegatggaaeaeaeggegatggaaeaeaeggegatggaaeaeaeggaaeaeaeggegatggaaeaeaeggegatggaaeaeaeggegaaaeaeaeggaaeaeaeae	i		agcacacggt	cacactgctt	ceggtagtea	acaaaccygc	adaccaycaa
ggcgtagcac caggcgttta agggcaccaa taactgcctt taagaaaatta tgcgcccgccc tgccactcat cgcagtactg tgtaattca ttaagcattc tgccgacatg gaagccatca cagacggcat ggcatcagca ccttgtcgcc ttgcgtataa tatttgccca tggtgaaaac gggggggaag aagttgtcca tattggccac gtttaaatca aaactggtga aactcaccca gggattggct gagacgaaaa acatattctc aatagggaaat tattgtgtaga aactgccgga aatcgccgga aatcgccgac ttgggaatacct cacgccacat cttgcgaata tattgtgtaga aactgccgga aatcgccgga ggattggct gagacgaaaa catattctc aataaaccct tatgtgtaga aactgccgga aatcgccgta ggtattcactc cacgccacat cttgcgaata tatgtgtaga aactgccgga agttgcca tggaaaacgg tgtatcactc cagagcgatg gtgaacacta ccagccacat cttgcgaata cacgccacat cacgccacat cttgcgaata cacgccacat cttgcgaata aactgctcgga agttgcca tggaaaacgg tgtaacacagg gtgaacacta ccagcccact cacgccacat cttgcgaata cacgccacat cacgccacat cttgcgaata acctgctcgaacacggatg gtgaacacta ccagccacat cacgccacat cacacacatacacacacacacacacacacaca			tagacataag	cggctattta	acgaccctgc	cccgaaccga	cyaccygytc
cgcccgccc tgccactcat cgcagtactg ttgtaattca ttaagcattc tgccgacatg gaagccatca cagacggcat gatgaacctg aatcgccagc ggcatcagca ccttgtcgcc ttgcgtataa tatttgccca tggtgaaaac gggggggaag aagttgtcca tattggccac gtttaaatca aaactggtga aactcaccca gggattggct gagacgaaaa acatattctc aataaaccct ttagggaaat aggccaggtt ttcaccgtaa cacgccacat cttgcgaata tatgtgtaga aactgccgga aatcgtcgtg gtattcactc cagagcgatg aaaacgtttc agtttgctca tggaaaacgg tgtaacaagg gtgaacacta tcccatatca ccagctcacc gtctttcatt gccatacg SEQ ID: 21 pKQ ATGGATCCGA GCTCGAGATC TGCAGCTGGT ACCATATGGG AATTCGAAGC TTGGGCCCGA ACAAAACTC ATCTCAGAAG AGGATCTGAA TAGCGCCGTC GACCATCATC ATCATCATCA TTGAGTTTAA ACCGTCTCCA GCTTGGCTGT TTTGGCGGAT GAGAGAAGAT TTTCAGCCTG ATACAGATTA AATCAGAACG CAGAAGCGGT CTGATAAAAC AGAATTTGCC TGGCGGCAGT AGCGCCGAT TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT			gaatttgctt	ccgaatttct	gccattcatc	tactattat	cacciacida
tgccgacatg gaagccatca cagacggcat gatgaacctg aatcgccagc ggcatcagca ccttgtcgcc ttgcgtataa tatttgcca tggtgaaaac ggggggcgaag aagttgtca tattggccac gtttaaatca aaactggtga aactcaccca gggattggct gagacgaaaa acatattctc aataaaccct ttagggaaat aggccaggtt ttcaccgtaa cacgccacat cttgcgaata tatgtgtaga aactgccgga aatcgtcgtg gtattcactc cagagcgatg aaacggttc agtttgctca tggaaaacgg gtgaacacta tcccatatca ccagctcacc gtctttcatt gccatacg SEQ ID: 21 pKQ ATGGATCCGA GCTCGAGATC TGCAGCTGGT ACCATATGGG AATTCGAAGC TTGGGCCCGA ACAAAAACTC ATCTCAGAAG AGGATCTGAA TAGCGCCGTC GACCATCATC ATCATCATCA TTGAGTTTAA ACGGTCTCCA GCTTGGCTGT TTTGGCCGGAT GAGAGAAGAT TTTCAGCCTG ATACAGATTA AATCAGAACG CAGAAGCGGT CTGATAAAAC AGAATTGCC TGGCGGCAGT AGCGCCGAT TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT			ggcgtagcac	caggcgttta	agggcaccaa	Laactgcctt	ttaagaatta
ggcatcagca ccttgtcgcc ttgcgtataa tatttgcca tggtgaaaac gggggcgaag aagttgtcaa tattggcac gtttaaatca aaactggtga aactcacca gggattggct gagacgaaaa acatattctc aataaaccct ttagggaaat aggccaggtt ttcaccgtaa cacgccacat cttgcgaata tatgtgtaga aactgccgga aatcgtcgtg gtattcactc cagagcgatg aaaacgtttc agtttgctca tggaaaacgg tgtaacaagg gtgaacacta tcccatatca ccagctcacc gtcttcatt gccatacg SEQ ID: 21 pKQ ATGGATCCGA GCTCGAGATC TGCAGCTGGT ACCATATGGG AATTCGAAGC TTGGGCCCGA ACAAAAACTC ATCATCAAGA AGGATCTGAA TAGCGCCGTC GACCATCATC ATCATCATCA TTGAGTTTAA ACGGTCTCCA GCTTGGCTGT TTTGGCGGAT GAGAGAAGAT TTTCAGCCTG ATACAGATTA AATCAGAACG CAGAAGCGGT CTGATAAAAC AGAATTTGCC TGGCGCAGT AGCGCCGTGG TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT			cgccccgccc	tgccactcat	cgcagtactg	utgtaattca	not agains
gggggcgaag aagttgtcca tattggccac gtttaaatca aaactggtga aactcaccca gggattggct gagacgaaaa acatattctc aataaaccct ttagggaaat aggccaggtt ttcaccgtaa cacgccacat cttgcgaata tatgtgtaga aactgccgga aatcgtcgtg gtattcactc cagagcgatg aaaacgtttc agtttgctca tggaaaacgg tgtaacaagg gtgaacacta tcccatatca ccagctcacc gtctttcatt gccatacg SEQ ID: 21 pKQ ATGGATCCGA GCTCGAGATC TGCAGCTGGT ACCATATGGG AATTCGAAGC TTGGGCCCGA ACAAAAACTC ATCTCAGAAG AGGATCTGAA TAGCGCCGTC GACCATCATC ATCATCATCA TTGAGTTTAA ACGGTCTCCA GCTTGGCTGT TTTGGCGGAT GAGAGAAGAT TTTCAGCCTG ATACAGATTA AATCAGAACG CAGAAGCGGT CTGATAAAAC AGAATTTGCC TGGCGGCAGT AGCGCCGTGG TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT	}		tgccgacatg	gaagccatca	cagacggcat	gatgaacctg	aacegeeage
aactcacca gggattggct gagacgaaa acatattete aataaaccet ttagggaat aggccaggtt ttcaccgtaa cacgccacat cttgcgaata tatgtgtaga aactgccgga aatcgtcgtg gtattcacte cagagcgatg aaaacgttte agtttgetea tggaaaacgg tgtaacaagg gtgaacacta tcccatatea ccagetcace gtettteatt gccatacg SEQ ID: 21 pKQ ATGGATCCGA GCTCGAGATC TGCAGCTGGT ACCATATGGG AATTCGAAGC TTGGGCCCGA ACAAAAACTC ATCTCAGAAG AGGATCTGAA TAGCGCCGTC GACCATCATC ATCATCATCA TTGAGTTTAA ACGGTCTCCA GCTTGGCTGT TTTGGCGGAT GAGAGAAGAT TTTCAGCCTG ATACAGATTA AATCAGAACG CAGAAGCGGT CTGATAAAAC AGAATTTGCC TGGCGGCAGT AGCGCCGTGG TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT			ggcatcagca	ccttgtcgcc	ttgcgtataa	tatttgccca	rggrgaaaac
ttagggaat aggccaggtt ttcaccgtaa cacgccacat cttgcgaata tatgtgtaga aactgccgga aatcgtcgtg gtattcactc cagagcgatg aaacgtttc agtttgctca tggaaaacgg tgtaacaagg gtgaacacta tcccatatca ccagctcacc gtctttcatt gccatacg SEQ ID: 21 pKQ ATGGATCCGA GCTCGAGATC TGCAGCTGGT ACCATATGGG AATTCGAAGC TTGGGCCCGA ACAAAAACTC ATCTCAGAAG AGGATCTGAA TAGCGCCGTC GACCATCATC ATCATCATCA TTGAGTTTAA ACGGTCTCCA GCTTGGCTGT TTTGGCGGAT GAGAAGAGAT TTTCAGCCTG ATACAGATTA AATCAGAACG CAGAAGCGGT CTGATAAAAC AGAATTTGCC TGGCGGCAGT AGCGCCGATG TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT			gggggcgaag	aagttgtcca	tattggccac	gtttaaatca	aaactggtga
tatgtgtaga aactgccgga aatcgtcgtg gtattcactc cagagcgatg aaacagtttc agtttgctca tggaaaacgg tgtaacaagg gtgaacacta tcccatatca ccagctcacc gtcttcatt gccatacg SEQ ID: 21 pKQ ATGGATCCGA GCTCGAGATC TGCAGCTGGT ACCATATGGG AATTCGAAGC TTGGGCCCGA ACAAAAACTC ATCTCAGAAG AGGATCTGAA TAGCGCCGTC GACCATCATC ATCATCATCA TTGAGTTTAA ACGGTCTCCA GCTTGGCTGT TTTGGCGGAT GAGAAGAGAT TTTCAGCCTG ATACAGATTA AATCAGAACG CAGAAGCGGT CTGATAAAAC AGAATTTGCC TGGCGGCAGT AGCGCCGATG TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT			aactcaccca	gggattggct	gagacgaaaa	acatattctc	aataaaccct
tatgtgtaga aactgccgga aatcgtcgtg gtattcactc cagagcgatg aaacagtttc agtttgctca tggaaaacgg tgtaacaagg gtgaacacta tcccatatca ccagctcacc gtcttcatt gccatacg SEQ ID: 21 pKQ ATGGATCCGA GCTCGAGATC TGCAGCTGGT ACCATATGGG AATTCGAAGC TTGGGCCCGA ACAAAAACTC ATCTCAGAAG AGGATCTGAA TAGCGCCGTC GACCATCATC ATCATCATCA TTGAGTTTAA ACGGTCTCCA GCTTGGCTGT TTTGGCGGAT GAGAAGAGAT TTTCAGCCTG ATACAGATTA AATCAGAACG CAGAAGCGGT CTGATAAAAC AGAATTTGCC TGGCGGCAGT AGCGCCGATG TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT			ttagggaaat	aggccaggtt	ttcaccgtaa	cacgccacat	cttgcgaata
SEQ ID: 21 pKQ ATGGATCCGA GCTCGAGATC TGCAGCTGGT ACCATATGGG AATTCGAAGC TTGGGCCCGA ACAAAAACTC ATCATCAAGA AGGATCTCAA TAGCGCCGTC GACCATCATC ATCATCATCA TTGAGTTTAA ACGGTCTCCA GCTTGGCTGT TTTGGCGGAT GAGAGAAGAT TTTCAGCCTG ATACAGATTA AATCAGAACG CAGAAGCGGT CTGATAAAAC AGAATTTGCC TGGCGGCAGT AGCGCGGTGG TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT	1		tatgtgtaga	aactgccgga	aatcgtcgtg	gtattcactc	cagagcgatg
SEQ ID: 21 pKQ ATGGATCCGA GCTCGAGATC TGCAGCTGGT ACCATATGGG AATTCGAAGC TTGGGCCCGA ACAAAAACTC ATCTCAGAAG AGGATCTGAA TAGCGCCGTC GACCATCATC ATCATCAT ATGAGTTTAA ACGGTCTCCA GCTTGGCTGT TTTGGCGGAT GAGAGAAGAT TTTCAGCCTG ATACAGATTA AATCAGAACG CAGAAGCGGT CTGATAAAAC AGAATTTGCC TGGCGCAGT AGCGCGGTGG TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT			aaaacgtttc	agtttgctca	tggaaaacgg	tgtaacaagg	gtgaacacta
SEQ ID: 21 pKQ ATGGATCCGA GCTCGAGATC TGCAGCTGGT ACCATATGGG AATTCGAAGC TTGGGCCCGA ACAAAAACTC ATCTCAGAAG AGGATCTGAA TAGCGCCGTC GACCATCATC ATCATCAT TTGAGTTTAA ACGGTCTCCA GCTTGGCTGT TTTGGCGGAT GAGAGAAGAT TTTCAGCCTG ATACAGATTA AATCAGAACG CAGAAGCGGT CTGATAAAAC AGAATTTGCC TGGCGCAGT AGCGCGGTGG TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT			tcccatatca	ccagctcacc	gtctttcatt	gccatacg	İ
TTGGGCCCGA ACAAAAACTC ATCTCAGAAG AGGATCTGAA TAGCGCCGTC GACCATCATC ATCATCATCA TTGAGTTTAA ACGGTCTCCA GCTTGGCTGT TTTGGCGGAT GAGAGAAGAT TTTCAGCCTG ATACAGATTA AATCAGAACG CAGAAGCGGT CTGATAAAAC AGAATTTGCC TGGCGGCAGT AGCGCGGTGG TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT					_		
TTGGGCCCGA ACAAAAACTC ATCTCAGAAG AGGATCTGAA TAGCGCCGTC GACCATCATC ATCATCATCA TTGAGTTTAA ACGGTCTCCA GCTTGGCTGT TTTGGCGGAT GAGAGAAGAT TTTCAGCCTG ATACAGATTA AATCAGAACG CAGAAGCGGT CTGATAAAAC AGAATTTGCC TGGCGGCAGT AGCGCGGTGG TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT	SEO ID: 21	nKO	ATGGATCCGA	GCTCGAGATC	TGCAGCTGGT	ACCATATGGG	AATTCGAAGC
GACCATCATC ATCATCATCA TTGAGTTTAA ACGGTCTCCA GCTTGGCTGT TTTGGCGGAT GAGAGAGAT TTTCAGCCTG ATACAGATTA AATCAGAACG CAGAAGCGGT CTGATAAAAC AGAATTTGCC TGGCGGCAGT AGCGCGGTGG TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT	322 22.21	hvz	TTGGGCCCGA	ACAAAAACTC	ATCTCAGAAG	AGGATCTGAA	TAGCGCCGTC
TTTGGCGGAT GAGAGAAGAT TTTCAGCCTG ATACAGATTA AATCAGAACG CAGAAGCGGT CTGATAAAAC AGAATTTGCC TGGCGGCAGT AGCGCGGTGG TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT			GACCATCATC	ATCATCATCA	TTGAGTTTAA	ACGGTCTCCA	GCTTGGCTGT
CAGAAGCGGT CTGATAAAAC AGAATTTGCC TGGCGGCAGT AGCGCGGTGG TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT			ተተተርርርር ይጥ	GAGAGAAGAT	TTTCAGCCTG	ATACAGATTA	AATCAGAACG
TCCCACCTGA CCCCATGCCG AACTCAGAAG TGAAACGCCG TAGCGCCGAT			CAGAAGCGGAI	CTCATAAAAC	AGAATTTCCC	TGGCGGCAGT	AGCGCGGTGG
GGTAGTGTGG GGTCTCCCCA TGCGAGAGTA GGGAACTGCC AGGCATCAAA			THURAGE GGT	CLOSTWOOL	AACTCAGAAG	TGAAACGCCG	TAGCGCCGAT
GGINGIGIGG GGICICCCCA IGCGAGAGIA GGGAACIGCC AGGCAICAAA			CCMACMOMO	CCCCAIGCCG	THE LEAGUED	CCCD ACTICCCC	ACCCATCAAA
			GGTAGTGTGG	GGTCTCCCCA	IGCGAGAGIA	GGGAACIGCC	1100C111C11F11

TAAAACGAAA GGCTCAGTCG AAAGACTGGG CCTTTCGTTT TATCTGTTGT TTGTCGGTGA ACGATATCTG CTTTTCTTCG CGAATtaatt ccgcttcgca ACATGTgagc aaaaggccag caaaaggcca ggaaccgtaa aaaggccgcg ttgctggcgt ttttccatag gctccgcccc cctgacgagc atcacaaaaa tcgacgctca aqtcaqaqqt qqcqaaaccc gacaggacta taaagatacc aggogtttcc ccctggaagc tccctcgtgc gctctcctgt tccgaccctg ccgcttaccg gatacctgtc cgcctttctc ccttcgggaa gcgtggcgct ttctcatagc tcacgctgta ggtatctcag ttcggtgtag gtcgttcgct ccaagctggg ctgtgtgcac gaaccccccg ttcagcccga ccgctgcgcc ttatccggta actatcgtct tgagtccaac ccggtaagac acgacttatc gccactggca gcagccactg gtaacaggat tagcagagcg aggtatgtag gcggtgctac agagttcttg aagtggtggc ctaactacgg ctacactaga aggacagtat ttggtatctg cgctctgctg aagccagtta ccttcggaaa aagagttggt agctcttgat ccggcaaaca aaccaccgct ggtagcggtg gtttttttgt ttgcaagcag cagattacgc gcagaaaaaa aggatctcaa gaagateett tgatetttte tacggggtet gacgeteagt ggaacgaaaa ctcacqttaa qqqattttqq TCATGAqttq tqtctcaaaa tctctqatqt tacattgcac aagataaaaa tatatcatca tgaacaataa aactgtctgc ttacataaac agtaatacaa ggggtgttat gagccatatt caacgggaaa cgtcttgctc gaggccgcga ttaaattcca acatggatgc tgatttatat gggtataaat gggctcgcga taatgtcggg caatcaggtg cgacaatcta tcgattgtat gggaagcccg atgcgccaga gttgtttctg aaacatggca aaggtagcgt tgccaatgat gttacagatg agatggtcag actaaactgg ctgacggaat ttatgcctct tccgaccatc aagcatttta tccgtactcc tgatgatgca tggttactca ccactgcgat ccccgggaaa acagcattcc aggtattaga agaatatcct gattcaggtg aaaatattgt tgatgcgctg gcagtgttcc tgcgccggtt gcattcgatt cctgtttgta attgtccttt taacagcgat cgcgtatttc gtctcgctca ggcgcaatca cgaatgaata acggtttggt tgatgcgagt gattttgatg acgagcgtaa tggctggcct qttqaacaaq tctqqaaaga aatgcataag cttttgccat tctcaccgga ttcagtcgtc actcatggtg atttctcact tgataacctt atttttgacg aggggaaatt aataggttgt attgatgttg gacgagtcgg aatcgcagac cgataccagg atcttgccat cctatggaac tgcctcggtg agttttctcc ttcattacag aaacgcttt ttcaaaaata tggtattgat aatcctgata tgaataaatt.gcagtttcat ttgatgctcg atgagttttt ctaatcagaa ttggttaatt ggttgtaaca ctggcagagc attacgctga cttgacggga cggcggcttt gttgaataaa tcgaactttt gctgagttga aggatcCTCG GGagttgtca gcctgtcccg cttataagat catacgccgt tatacGTTGT TTACGCTTTG AGGAATTAAC C